

## On Learning Discrete-Time Fractional-Order Dynamical Systems

Chatterjee, Sarthak; Gonçalves Melo Pequito, S.D.

**DOI**

[10.23919/ACC53348.2022.9867773](https://doi.org/10.23919/ACC53348.2022.9867773)

**Publication date**

2022

**Document Version**

Final published version

**Published in**

Proceedings of the American Control Conference (ACC 2022)

**Citation (APA)**

Chatterjee, S., & Gonçalves Melo Pequito, S. D. (2022). On Learning Discrete-Time Fractional-Order Dynamical Systems. In *Proceedings of the American Control Conference (ACC 2022)* (pp. 4335-4340). IEEE. <https://doi.org/10.23919/ACC53348.2022.9867773>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

***Green Open Access added to TU Delft Institutional Repository***

***'You share, we take care!' - Taverne project***

***<https://www.openaccess.nl/en/you-share-we-take-care>***

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

# On Learning Discrete-Time Fractional-Order Dynamical Systems

Sarthak Chatterjee<sup>†</sup>

Sérgio Pequito<sup>‡</sup>

**Abstract**—Discrete-time fractional-order dynamical systems (DT-FODS) have found innumerable applications in the context of modeling spatiotemporal behaviors associated with long-term memory. Applications include neurophysiological signals such as electroencephalogram (EEG) and electrocorticogram (ECoG). Although learning the spatiotemporal parameters of DT-FODS is not a new problem, when dealing with neurophysiological signals we need to guarantee performance standards. Therefore, we need to understand the trade-offs between sample complexity and estimation accuracy of the system parameters. Simply speaking, we need to address the question of how many measurements we need to collect to identify the system parameters up to an uncertainty level. In this paper, we address the problem of identifying the spatial and temporal parameters of DT-FODS. The main result is the first result on non-asymptotic finite-sample complexity guarantees of identifying DT-FODS. Finally, we provide evidence of the efficacy of our method in the context of forecasting real-life intracranial EEG time series collected from patients undergoing epileptic seizures.

## I. INTRODUCTION

Learning a model (or system identification) is the problem of estimating the parameters of a dynamical system given an input-output time series of the trajectories that have been generated by the former. In many problem settings in control theory, time series analysis, reinforcement learning, and econometrics, estimating the parameters of a system from its input-output behavior is an important problem, especially in the absence of analytical tools to find the underlying dynamics. Extensive attention has been paid to the topic, with [1] and [2] being comprehensive references on the same.

Lately, there has been a renewed interest in investigating finite-sample guarantees for problems in classical system identification and control theory from the lens of statistical learning theory. On the one hand, significant work has been done on developing finite-time regret guarantees for the linear-quadratic regulator when the latter is trying to control a system with unknown dynamics [3]–[8]. Alternatively, finite-sample guarantees have also been derived in the context of high probability bounds with respect to the norm of the estimation error of the system's parameters in a wide variety of problem settings [9]–[22].

Nonetheless, most of the work focuses on cases where there is a Markovian dependence of the current state of a system on just the previous state, which is insufficient in describing the long-term behavior of the aforementioned

systems. Yet, systems that we often encounter in real-life demonstrate phenomena such as hysteresis and long-term memory, in which the current system state is dependent on a combination of several past states or the entire gamut of states seen so far in time.

In particular, fractional-order dynamical systems (FODS) have attracted a lot of interest in recent years for being able to successfully model a wide variety of dynamical system behaviors, chief among them being (nonexponential) power-law decay in the dependence of the current state on past states, systems exhibiting long-term memory or fractal properties, or dynamics where there are adaptations across multiple time scales [23]–[27]. FODS have been used in domains as disparate as biological swarms [28], chaotic systems [29], gaseous dynamics [30], and cyber-physical systems [31], to mention a few. In the context of this work, we emphasize the applications to neuromodulation as well as modeling neurophysiological signals such as electroencephalogram (EEG) or electrocorticogram (ECoG) [32]–[34].

In this paper, we will investigate the problem of learning the spatial and temporal parameters of a discrete-time fractional-order dynamical system (DT-FODS). We provide finite-sample and iteration complexity guarantees for the same, and we further illustrate the working of our proposed approach on real-life intracranial EEG (iEEG) data collected from patients undergoing epileptic seizures. In the context of using model-based approaches for neurostimulation (such as to perform epileptic seizure suppression [33]), robust guarantees for the iteration and finite-sample complexities of learning the model parameters is of paramount importance. Additionally, using approaches such as model predictive control (MPC) in order to develop neurostimulation strategies often require certifications on the minimum amount of data required in order to learn robust models.

The rest of the paper is structured as follows. Section II introduces DT-FODS and the problem that we seek to address in this paper. Section III outlines an approach to learn the parameters of a DT-FODS as well as the main results in terms of the finite-sample and iteration complexity guarantees for the same. Section IV demonstrates the working of our approach on real-life iEEG data. Section V concludes the paper.

## II. PROBLEM STATEMENT

We consider a linear discrete-time fractional-order dynamical system described as follows

$$\begin{aligned}\Delta^\alpha x[k+1] &= Ax[k] + w[k] \\ x[0] &= x_0,\end{aligned}\tag{1}$$

<sup>†</sup>Sarthak Chatterjee is with the department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA sarthak.chatterjee92@gmail.com

<sup>‡</sup>Sérgio Pequito is with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands Sergio.Pequito@tudelft.nl

where  $x[k] \in \mathbb{R}^n$  is the *state* for time step  $k \in \mathbb{N}$  and  $A \in \mathbb{R}^{n \times n}$  is the *system matrix*. The *process noise*  $w[k]$  is assumed to be an i.i.d. additive white noise process originating from the standard normal distribution, i.e.,  $w[k] \sim \mathcal{N}(0, \sigma^2 I_n)$ . The above system model is similar to a classic discrete-time linear time-invariant model except for the inclusion of the Grünwald-Letnikov fractional derivative, whose expansion and discretization for the  $i$ -th state,  $1 \leq i \leq n$ , can be written as

$$\Delta^{\alpha_i} x_i[k] = \sum_{j=0}^k \psi(\alpha_i, j) x_i[k-j], \quad (2)$$

where  $\alpha_i$  is the fractional order corresponding to state  $i$  (with  $\alpha_i > 0$  for all  $i$ ) and

$$\psi(\alpha_i, j) = \frac{\Gamma(j - \alpha_i)}{\Gamma(-\alpha_i)\Gamma(j+1)}, \quad (3)$$

with  $\Gamma(\cdot)$  being the gamma function defined by  $\Gamma(z) = \int_0^\infty s^{z-1} e^{-s} ds$  for all complex numbers  $z$  with  $\Re(z) > 0$  [35].

Given the above ingredients, we seek to solve the following problem in this paper.

**Problem 1.** Consider the discrete-time fractional-order dynamical system model given in (1) with  $w[k]$  being an additive white Gaussian process noise. We seek to provide non-asymptotic finite-sample as well as iteration complexity guarantees of identifying the temporal and spatial system parameters of (1) ( $\{\alpha_i\}_{i=1}^n$  and  $A$ , respectively) using an observed trajectory of the states.

### III. LEARNING DISCRETE-TIME FRACTIONAL-ORDER DYNAMICAL SYSTEMS

In this section, we establish in a sequential manner, a bi-level iterative scheme to learn the spatial and temporal components of a DT-FODS. We first start with some fundamentals regarding DT-FODS.

#### A. Preliminaries

Let us first review some essential theory for fractional-order systems, including an approximation of (1) as an LTI system. Using the expansion of the Grünwald-Letnikov derivative in (2), we have

$$\begin{aligned} \Delta^\alpha x[k] &= \begin{bmatrix} \Delta^{\alpha_1} x_1[k] \\ \vdots \\ \Delta^{\alpha_n} x_n[k] \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^k \psi(\alpha_1, j) x_1[k-j] \\ \vdots \\ \sum_{j=0}^k \psi(\alpha_n, j) x_n[k-j] \end{bmatrix} \\ &= \sum_{j=0}^k \underbrace{\begin{bmatrix} \alpha_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha_n \end{bmatrix}}_{D(\alpha, j)} \begin{bmatrix} x_1[k-j] \\ \vdots \\ x_n[k-j] \end{bmatrix} \\ &= \sum_{j=0}^k D(\alpha, j) x[k-j]. \end{aligned} \quad (4)$$

The above formulation distinctly highlights one of the main peculiarities of DT-FODS in that the fractional derivative

$\Delta^\alpha x[k]$  is a weighted linear combination of not just the previous state but of every single state up to the current one, with the weights given by (3) following a power-law decay.

Plugging (4) into the DT-FODS formulation (1), we have

$$\sum_{j=0}^{k+1} D(\alpha, j) x[k+1-j] = Ax[k] + w[k], \quad (5)$$

or, equivalently,

$$D(\alpha, 0) x[k+1] = - \sum_{j=1}^{k+1} D(\alpha, j) x[k+1-j] + Ax[k] + w[k], \quad (6)$$

which leads to

$$x[k+1] = - \sum_{j=0}^k D(\alpha, j+1) x[k-j] + Ax[k] + w[k], \quad (7)$$

since  $D(\alpha, 0) = I_n$ . Alternatively, (7) can be written as

$$\begin{aligned} x[k+1] &= \sum_{j=0}^k A_j x[k-j] + w[k] \\ x[0] &= x_0, \end{aligned} \quad (8)$$

where

$$A_j = \begin{cases} A - \text{diag}(\alpha_1, \dots, \alpha_n) & \text{if } j = 0 \\ -D(\alpha, j+1) & \text{if } j \geq 1 \end{cases}. \quad (9)$$

In the subsequent discussion, we will consider a truncation of the last  $p$  temporal components of (1). Defining

$$\tilde{x}[k] = \begin{bmatrix} x[k] \\ x[k-1] \\ \vdots \\ x[k-p+1] \end{bmatrix} \quad (10)$$

as the *augmented* state vector and assuming that the system is *causal*, i.e., the state and disturbances are all considered to be zero before the initial time (i.e.,  $x[k] = 0$  and  $w[k] = 0$  for all  $k < 0$ ), we have

$$\begin{aligned} \tilde{x}[k+1] &= \underbrace{\begin{bmatrix} A_0 & \dots & A_{p-2} & A_{p-1} \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix}}_{\tilde{A}} \tilde{x}[k] + \underbrace{\begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\tilde{B}^w} w[k] \\ &= \tilde{A} \tilde{x}[k] + \tilde{B}^w w[k], \end{aligned} \quad (11)$$

for all  $k \geq 0$ . Note that (11) is an LTI system model, which we refer to as the *p-augmented LTI approximation* of (1).

#### B. Two-level iterative bisection scheme to identify the parameters of a DT-FODS

Having established the  $p$ -augmented LTI approximation of a DT-FODS in (11), we will now use a two-level iterative bisection-like approach to identify the spatial and temporal parameters of the DT-FODS in (1). In particular, we start by noting the fact that for the Grünwald-Letnikov definition of

the fractional derivative provided in (2),  $\alpha_i = 1$  and  $\alpha_i = -1$  can be interpreted, respectively, to be the discretized version of the derivative and the integral for  $1 \leq i \leq n$ , as defined in the sense of ordinary calculus.

In order to proceed with a bisection-like approach to identify  $\{\alpha_i\}_{i=1}^n$  and  $\tilde{A}$ , we first fix the endpoints of the search space for  $\alpha_i$  to be  $\underline{\alpha}_i = -1$  and  $\overline{\alpha}_i = 1$  for  $1 \leq i \leq n$ . We also calculate the value of  $\alpha_{c,i} = (\underline{\alpha}_i + \overline{\alpha}_i)/2$ . Now, given the values of  $\underline{\alpha}_i$ ,  $\overline{\alpha}_i$ , and  $\alpha_{c,i}$ , we calculate using the ordinary least squares (OLS) technique described in Section III-C below, the row vectors  $\tilde{a}_i$ ,  $\tilde{\bar{a}}_i$ , and  $\tilde{a}_{c,i}$ , respectively, that guide the evolution of the states in the  $p$ -augmented LTI approximation

$$\tilde{x}_i[k+1] = \tilde{a}_i \tilde{x}_i[k] + \tilde{b}_i^w w_i[k], \quad (12)$$

where  $\tilde{a}_i = \tilde{a}_i$  when  $\alpha_i = \underline{\alpha}_i$ ,  $\tilde{a}_i = \tilde{\bar{a}}_i$  when  $\alpha_i = \overline{\alpha}_i$ , and  $\tilde{a}_i = \tilde{a}_{c,i}$  when  $\alpha_i = \alpha_{c,i}$  with  $\tilde{b}_i^w$  being obtained by extracting the  $i$ -th row of  $\tilde{B}^w$  for  $1 \leq i \leq n$ .

Next, we propagate the dynamics according to the obtained values of the parameters  $\tilde{a}_i$  and calculate the mean squared error (MSE) between the states obtained as a result of the estimated  $\tilde{a}_i$ 's and the observed states. If the MSE is lesser corresponding to the  $\underline{\alpha}_i$  case, then we set  $\overline{\alpha}_i = \alpha_{c,i}$ . If the MSE is lesser corresponding to the  $\overline{\alpha}_i$  case, then we set  $\underline{\alpha}_i = \alpha_{c,i}$ . This approach is repeated until  $|\overline{\alpha}_i - \underline{\alpha}_i|$  is lesser than a certain pre-specified tolerance  $\varepsilon$ . Algorithm 1 summarizes the procedure of determining the spatial and temporal components of a DT-FODS using the two-level iterative bisection-like approach we have outlined above.

Empirically, numerical and experimental evidence suggests that the computation of the temporal parameters of a DT-FODS using, e.g., a wavelet-like technique described in [36], does not directly depend on the number of samples or observations used for the aforementioned estimation procedure. Empirical evidence suggests that a small number of samples (usually 30 to 100) suffice in order to compute  $\{\alpha_i\}_{i=1}^n$ . Therefore, for the estimation of the temporal components of a DT-FODS, we specify the iteration complexity of the bisection-like process and then, in Section III-D, we investigate the finite-sample complexity of computing the spatial parameters using a least squares approach.

Apropos the above discussion, in the next result, we certify the iteration complexity of the bisection method to find the spatial and temporal parameters of a DT-FODS.

**Theorem 1.** *The bisection-based technique detailed above to find the temporal components of a DT-FODS is minmax optimal and the number  $\nu$  of iterations needed in order to achieve a certain specified tolerance  $\varepsilon$  when this technique is used is bounded above by*

$$\nu \leq \left\lceil \log_2 \left( \frac{2}{\varepsilon} \right) \right\rceil. \quad (13)$$

*Proof.* See [37] for a proof. ■

---

#### Algorithm 1 Learning the parameters of a DT-FODS

---

- 1: **for**  $i = 1$  to  $n$  **do**
  - 2:   Initialize  $\underline{\alpha}_i = -1$ ,  $\overline{\alpha}_i = 1$ , and tolerance  $\varepsilon$ .
  - 3:   Calculate  $\alpha_{c,i} = (\underline{\alpha}_i + \overline{\alpha}_i)/2$ .
  - 4:   Given the above values of  $\underline{\alpha}_i$ ,  $\overline{\alpha}_i$ , and  $\alpha_{c,i}$ , find, using the ordinary least squares (OLS) method, the row vectors  $\tilde{a}_i$ ,  $\tilde{\bar{a}}_i$ , and  $\tilde{a}_{c,i}$ , respectively, that guide the evolution of the states in the  $p$ -augmented LTI approximation  $\tilde{x}_i[k+1] = \tilde{a}_i \tilde{x}_i[k] + \tilde{b}_i^w w_i[k]$ .
  - 5:   Propagate the dynamics according to the obtained OLS estimates and calculate the mean squared error (MSE) between the propagated states and the observed state trajectory.
  - 6:   **if** MSE is lesser for the  $\underline{\alpha}_i$  case **then**
  - 7:     Set  $\overline{\alpha}_i = \alpha_{c,i}$ .
  - 8:   **else if** MSE is lesser for the  $\overline{\alpha}_i$  case **then**
  - 9:     Set  $\underline{\alpha}_i = \alpha_{c,i}$ .
  - 10:   **end if**
  - 11:   Terminate if  $|\overline{\alpha}_i - \underline{\alpha}_i| < \varepsilon$ , else return to step 3.
  - 12: **end for**
- 

#### C. The ordinary least squares (OLS) method to identify the spatial parameters of a DT-FODS

Having outlined the details of the bi-level iterative bisection-like scheme to identify the spatial and temporal components of a DT-FODS, we now delve into the problem of identifying the spatial parameters using a least squares-like approach. We start with the  $p$ -augmented LTI model of (11), i.e.,

$$\tilde{x}[k+1] = \tilde{A}\tilde{x}[k] + \tilde{B}^w w[k]. \quad (14)$$

The OLS method then outputs the matrix  $\tilde{A}[K]$  as the solution of the following optimization problem

$$\tilde{A}[K] := \underset{\tilde{A} \in \mathbb{R}^{d \times d}}{\operatorname{argmin}} \sum_{k=1}^K \frac{1}{2} \|\tilde{x}[k+1] - \tilde{A}\tilde{x}[k]\|_2^2, \quad (15)$$

by observing the state trajectory of (11), i.e.,  $\{x[0], x[1], \dots, x[K+1]\}$ , and the process noise  $w[k]$  being i.i.d. zero-mean Gaussian.

#### D. Finite-sample complexity guarantees for OLS system identification of the spatial parameters of a DT-FODS

Prior to characterizing the sample complexity of the OLS method for the  $p$ -augmented LTI approximation of the DT-FODS, we define a few quantities of interest.

**Definition 1.** The *finite-time controllability Gramian* of the approximated system (11),  $W_t$ , is defined by

$$W_t := \sum_{j=0}^{t-1} \tilde{A}^j (\tilde{A}^j)^\top. \quad (16)$$

Intuitively, the controllability Gramian gives a quantitative measure of how much the system is excited when induced by the process noise  $w[k]$  acting as an input to the system.

**Definition 2.** Given a symmetric matrix  $A \in \mathbb{R}^{d \times d}$ , we define  $\lambda_{\max}(A)$  and  $\lambda_{\min}(A)$  to denote, respectively, the maximum and minimum eigenvalues of the matrix  $A$ .

**Definition 3.** For any square matrix  $A \in \mathbb{R}^{d \times d}$ , the *spectral radius* of the matrix  $A$ ,  $\rho(A)$ , is given by the largest absolute value of its eigenvalues.

**Definition 4.** The *operator norm* of a matrix is denoted by  $\|\cdot\|_{\text{op}}$ .

We then have our first result that characterizes the sample complexity of the above OLS method for the DT-FODS approximation.

**Theorem 2** ([17]). Fix  $\delta \in (0, 1/2)$  and consider the  $p$ -augmented system in (11), where  $\tilde{A} \in \mathbb{R}^{d \times d}$  is a marginally stable matrix (i.e.,  $\rho(\tilde{A}) \leq 1$ ) and  $w[k] \sim \mathcal{N}(0, \sigma^2 I)$ . Then, there exist universal constants  $c, C > 0$  such that,

$$\mathbb{P} \left[ \left\| \tilde{A}[K] - \tilde{A} \right\|_{\text{op}} \leq \frac{C}{\sqrt{K \lambda_{\min}(W_k)}} \times \sqrt{d \log \left( \frac{d}{\delta} \right) + \log \det(W_K W_k^{-1})} \right] \geq 1 - \delta, \quad (17)$$

for any  $k$  such that

$$\frac{K}{k} \geq c \left( d \log \left( \frac{d}{\delta} \right) + \log \det(W_K W_k^{-1}) \right) \quad (18)$$

holds.

*Proof.* The proof of the theorem is a consequence of Theorem 2.4 in [17]. In general, if we consider the time series  $(X[t], Y[t])_{t \geq 1}$ , where  $Y[t] = A_* X[t] + \eta[t]$ , with  $Y[t], \eta[t] \in \mathbb{R}^n, X[t] \in \mathbb{R}^d$ , and  $A_* \in \mathbb{R}^{n \times d}$ , the OLS estimate is obtained by solving the optimization problem

$$\hat{A}(T) := \underset{A \in \mathbb{R}^{n \times d}}{\operatorname{argmin}} \sum_{t=1}^T \frac{1}{2} \|Y[t] - AX[t]\|_2^2, \quad (19)$$

where  $T$  is the observation horizon. We consider  $\mathcal{F}_t := \sigma(\eta[0], \eta[1], \dots, \eta[t], X[1], \dots, X[t])$  to be a filtration generated by the states and the noise, and that  $\eta[t]|\mathcal{F}_{t-1}$  to be zero-mean and  $\sigma^2$ -sub-Gaussian (or, in other words, sub-Gaussian with a variance proxy  $\sigma^2$ ), i.e.,  $\mathbb{E}[\exp(\lambda \eta[t]|\mathcal{F}_t)] \leq \exp(\lambda^2 \sigma^2 / 2)$  for all  $\lambda \in \mathbb{R}$ .

If we fix  $\epsilon, \delta \in (0, 1)$ ,  $T \in \mathbb{N}$ , and there exist  $W_{\text{sb}}$  and  $\bar{W}$  such that  $0 \prec W_{\text{sb}} \preceq \bar{W}$  (where the  $W$ 's are the controllability Gramians defined as in Definition 1), then for a random sequence  $(X[t], Y[t])_{t \geq 1} \in (\mathbb{R}^d \times \mathbb{R}^n)^T$  that satisfies

- (a)  $Y[t] = A_* X[t] + \eta[t]$  with  $\eta[t]|\mathcal{F}_t$  being zero-mean and  $\sigma^2$ -sub-Gaussian,
- (b)  $\{X[1], \dots, X[T]\}$  satisfies the  $(k, W_{\text{sb}}, p)$ -martingale small-ball condition (Definition 2.1 in [17]), and,
- (c)  $\mathbb{P} \left[ \sum_{t=1}^T X[t] X[t]^T \not\preceq T \bar{W} \right] \leq \delta,$

we have that if

$$T \geq \frac{10k}{p^2} \left( \log \left( \frac{1}{\delta} \right) + 2d \log \left( \frac{10}{p} \right) + \log \det(\bar{W} W_{\text{sb}}^{-1}) \right), \quad (20)$$

then

$$\mathbb{P} \left[ \left\| \hat{A}(T) - A_* \right\|_{\text{op}} \leq \frac{90\sigma}{p} \sqrt{\frac{n + d \log \frac{10}{p} + \log \det \bar{W} W_{\text{sb}}^{-1} + \log \frac{1}{\delta}}{T \lambda_{\min}(W_{\text{sb}})}} \right] \geq 1 - 3\delta. \quad (21)$$

Further, for the specific case considered in this paper, we know that the process noise satisfies the  $\sigma^2$ -sub-Gaussian condition. The intuition behind this follows easily from the fact that if a random variable  $X$  has the distribution  $\mathcal{N}(0, \sigma^2)$ , then  $\mathbb{E}(\exp(\lambda X)) = \exp(\lambda^2 \sigma^2 / 2)$  for all  $\lambda \in \mathbb{R}$ , and thus  $X$  is  $\sigma^2$ -sub-Gaussian, with a similar argument holding for Gaussian random vectors as well. We further note that the observed length of our state trajectory is  $K$ . Hence, we can write

$$\begin{aligned} & \mathbb{P} \left[ \mathbf{X}^T \mathbf{X} \not\preceq \frac{d\sigma^2}{\delta} K W_K \right] \\ &= \mathbb{P} \left[ \lambda_{\max} \left( (K W_K)^{-1/2} \mathbf{X}^T \mathbf{X} (K W_K)^{-1/2} \right) \geq \frac{d\sigma^2}{\delta} \right] \\ &\stackrel{(\Delta)}{\leq} \frac{\delta}{d\sigma^2} \cdot \mathbb{E} \left[ \lambda_{\max} \left( (K W_K)^{-1/2} \mathbf{X}^T \mathbf{X} (K W_K)^{-1/2} \right) \right] \\ &\leq \frac{\delta}{d\sigma^2} \cdot \mathbb{E} \left[ \operatorname{tr} \left( (K W_K)^{-1/2} \mathbf{X}^T \mathbf{X} (K W_K)^{-1/2} \right) \right] \\ &\stackrel{(\nabla)}{\leq} \delta, \end{aligned} \quad (22)$$

where the rows of  $\mathbf{X} \in \mathbb{R}^{K \times d}$  are formed using  $\tilde{x}[k]$  for each  $k$ . The inequality  $(\Delta)$  is a consequence of Markov's inequality and the inequality  $(\nabla)$  is a consequence of  $\mathbb{E}[\mathbf{X}^T \mathbf{X}] = \sigma^2 \sum_{j=1}^K W_j \preceq \sigma^2 K W_K$  and the linearity of the trace operator. The proof of the theorem follows from setting  $\bar{W} = \frac{d\sigma^2}{\delta} W_K$ , the equality  $\log \det \left( \left( \frac{d}{\delta} \sigma^2 W_K \right) (\sigma^2 W_{\lfloor k/2 \rfloor})^{-1} \right) = d \log(d/\delta) + \log \det(W_K W_{\lfloor k/2 \rfloor}^{-1})$ , and the fact that  $\tilde{x}[k]$  satisfies the  $(k, \sigma^2 W_{\lfloor k/2 \rfloor}, \frac{3}{20})$ -block martingale small-ball condition (Proposition 3.1 of [17]). ■

**Remark 1.** We note here that although the operator norm parameter estimation error in (17) is stated in terms of  $\tilde{A}$ , the operator norm errors associated with the matrices  $A_0, A_1, \dots, A_{p-1}$  are strictly lesser compared to  $\left\| \tilde{A}[K] - \tilde{A} \right\|_{\text{op}}$ , since  $A_0, A_1, \dots, A_{p-1}$  are submatrices of  $\tilde{A}$ , and for any operator norm, the operator norm of a submatrix is upper bounded by one of the whole matrix (see Lemma A.9 of [38] for a proof).

**Remark 2.** A finite-sample complexity bound similar to the one presented in Theorem 2 can also be derived when we consider the ordinary least squares identification of the

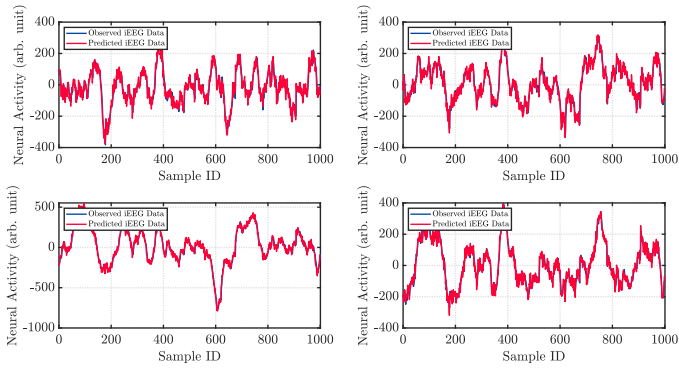


Fig. 1. Performance of our approach on real-life intracranial EEG data.

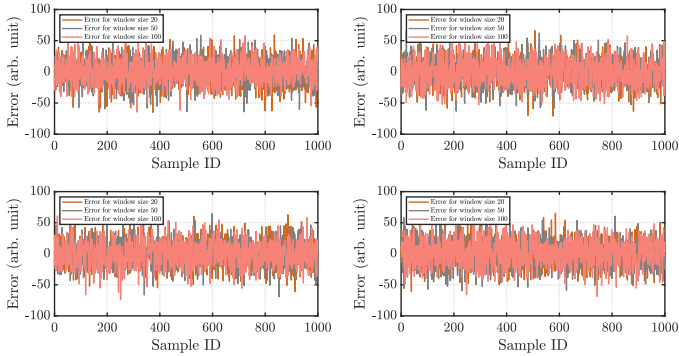


Fig. 2. Variation of the error of the least squares prediction with respect to the observed data, with varying window sizes in the least squares optimization problems.

spatial parameters of a DT-FODS with inputs. In the context of our work, an external input may correspond to an exogenous electrical stimulus or an actuation or control signal. Therefore, within the purview of epileptic seizure mitigation using intracranial EEG data, the objective of the former is to suppress the overall length or duration of an epileptic seizure, thus steering the state of the neurophysiological system in consideration away from seizure-like activity using a control strategy like model predictive control [33].

#### IV. SIMULATION RESULTS ON REAL-LIFE INTRACRANIAL EEG DATA

We present some preliminary results regarding the performance of the above approach on DT-FODS. Specifically, we use 1000 noisy measurements taken from 4 channels of an intracranial electroencephalographic (iEEG) signal which records the brain activity of subjects undergoing epileptic seizures. The signals were recorded and digitized at a sampling rate of 512 Hz at the Hospital of the University of Pennsylvania, Philadelphia, PA. Subdural grid and strip electrodes were placed at specialized locations (dictated by a multidisciplinary team of neurologists, neurosurgeons, and a radiologist), with the electrodes themselves consisting of linear and two-dimensional arrays spanning 2.3 mm in diameter and having an inter-contact spacing of 10 mm [39], [40]. All of the least squares optimization problems are solved using CVX [41], [42] with the aid of a window-based approach

using a finite subset of the entire range of measurements. This is done because the time series under consideration is nonlinear, and it is not possible to characterize the entire gamut of measurements using very few parameters. Figure 1 shows the performance of our method on the above data. Additionally, we also show in Figure 2 the variation of the error of the least squares predictions with respect to the observed data, with varying window sizes in the least squares optimization problems. We see that the identified system parameters are able to predict the system states fairly closely, thus demonstrating that our approach can be used to learn the system parameters of a DT-FODS.

#### V. CONCLUSIONS AND FUTURE WORK

In this paper, we present a framework that enables us to learn the spatial and temporal parameters of a DT-FODS. We provide non-asymptotic finite-sample complexity guarantees for the identification of the spatial components of autonomous and actuated DT-FODS using least squares as well as iteration complexity guarantees for the identification of the temporal components of a DT-FODS. We also show the efficacy of our proposed approach in identifying the system parameters of a DT-FODS that models iEEG signals characterizing the behavior of real-life subjects undergoing epileptic seizures.

There are a number of interesting directions which can be taken from here. A particular case that piques our interest is system identification using the Expectation-Maximization (EM) algorithm. Although approaches using the EM algorithm for linear [43] as well as nonlinear system identification [44] have existed in the literature for a while now, one immediately notices that there is a long-standing problem in characterizing theoretical robustness guarantees for the same. We are inspired by the preliminary analyses of finite-sample robustness guarantees for EM in [45]–[47], and, therefore, we wish to characterize the sample complexity in identifying LTI as well as fractional-order systems using the EM algorithm. Furthermore, the system identification of fractional-order systems is an extremely under-explored field in general, with a lack of a systematic and unified theory, with some preliminary approaches utilizing wavelets [36], frequency-domain techniques [48], [49], or a sequential combination of wavelets and EM [50]. Future work will focus on chalking out a general theory of identifying certain classes of fractional-order systems.

#### REFERENCES

- [1] L. Ljung, Ed., *System Identification (2nd Ed.): Theory for the User*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.
- [2] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall, Inc., 1988.
- [3] Y. Abbasi-Yadkori and C. Szepesvári, “Regret bounds for the adaptive control of linear quadratic systems,” in *Proceedings of the Conference on Learning Theory*. JMLR Workshop and Conference Proceedings, 2011, pp. 1–26.
- [4] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, “Regret bounds for robust adaptive control of the linear quadratic regulator,” in *Advances in Neural Information Processing Systems*, 2018, pp. 4192–4201.
- [5] —, “On the sample complexity of the linear quadratic regulator,” *Foundations of Computational Mathematics*, pp. 1–47, 2019.

- [6] S. Dean, S. Tu, N. Matni, and B. Recht, "Safely learning to control the constrained linear quadratic regulator," in *Proceedings of the 2019 American Control Conference (ACC)*. IEEE, 2019, pp. 5582–5588.
- [7] H. Mania, S. Tu, and B. Recht, "Certainty equivalence is efficient for linear quadratic control," *arXiv preprint arXiv:1902.07826*, 2019.
- [8] A. Cohen, T. Koren, and Y. Mansour, "Learning linear-quadratic regulators efficiently with only  $\sqrt{T}$  regret," in *Proceedings of the International Conference on Machine Learning*. PMLR, 2019, pp. 1300–1309.
- [9] S. Tu, R. Boczar, A. Packard, and B. Recht, "Non-asymptotic analysis of robust control from coarse-grained identification," *arXiv preprint arXiv:1707.04791*, 2017.
- [10] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Finite time identification in unstable linear systems," *Automatica*, vol. 96, pp. 342–353, 2018.
- [11] E. Hazan, H. Lee, K. Singh, C. Zhang, and Y. Zhang, "Spectral filtering for general linear dynamical systems," in *Advances in Neural Information Processing Systems*, 2018, pp. 4639–4648.
- [12] M. Hardt, T. Ma, and B. Recht, "Gradient descent learns linear dynamical systems," *Journal of Machine Learning Research*, vol. 19, no. 1, pp. 1025–1068, 2018.
- [13] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," in *Proceedings of the International Conference on Machine Learning*. PMLR, 2019, pp. 5610–5618.
- [14] T. Sarkar, A. Rakhlin, and M. A. Dahleh, "Finite time LTI system identification," *Journal of Machine Learning Research*, vol. 22, no. 26, pp. 1–61, 2021.
- [15] S. Oymak and N. Ozay, "Non-asymptotic identification of LTI systems from a single trajectory," in *Proceedings of the 2019 American Control Conference (ACC)*. IEEE, 2019, pp. 5655–5661.
- [16] A. Tsiamis and G. J. Pappas, "Finite sample analysis of stochastic system identification," in *Proceedings of the 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 3648–3654.
- [17] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Proceedings of the Conference on Learning Theory*. PMLR, 2018, pp. 439–473.
- [18] M. Simchowitz, R. Boczar, and B. Recht, "Learning linear dynamical systems with semi-parametric least squares," in *Proceedings of the Conference on Learning Theory*. PMLR, 2019, pp. 2714–2802.
- [19] A. Wagenmaker and K. Jamieson, "Active learning for identification of linear dynamical systems," in *Proceedings of the Conference on Learning Theory*. PMLR, 2020, pp. 3487–3582.
- [20] Y. Jedra and A. Proutiere, "Sample complexity lower bounds for linear system identification," in *Proceedings of the 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 2676–2681.
- [21] —, "Finite-time identification of stable linear systems optimality of the least-squares estimator," in *Proceedings of the 59th Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 996–1001.
- [22] Y. Zheng and N. Li, "Non-asymptotic identification of linear dynamical systems using multiple trajectories," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1693–1698, 2021.
- [23] F. C. Moon, *Chaotic and Fractal Dynamics: Introduction for Applied Scientists and Engineers*. John Wiley & Sons, 2008.
- [24] B. N. Lundstrom, M. H. Higgs, W. J. Spain, and A. L. Fairhall, "Fractional differentiation by neocortical pyramidal neurons," *Nature Neuroscience*, vol. 11, no. 11, p. 1335, 2008.
- [25] G. Werner, "Fractals in the nervous system: Conceptual implications for theoretical neuroscience," *Frontiers in Physiology*, vol. 1, p. 15, 2010.
- [26] S. Thurner, C. Windischberger, E. Moser, P. Walla, and M. Barth, "Scaling laws and persistence in human brain activity," *Physica A: Statistical Mechanics and its Applications*, vol. 326, no. 3–4, pp. 511–521, 2003.
- [27] M. C. Teich, C. Heneghan, S. B. Lowen, T. Ozaki, and E. Kaplan, "Fractal character of the neural spike train in the visual system of the cat," *Journal of the Optical Society of America*, vol. 14, no. 3, pp. 529–546, 1997.
- [28] B. J. West, M. Turla, and P. Grigolini, *Networks of Echoes: Imitation, Innovation and Invisible Leaders*. Springer Science & Business Media, 2014.
- [29] I. Petráš, "Fractional-order chaotic systems," in *Fractional-Order Nonlinear Systems*. Springer, 2011, pp. 103–184.
- [30] W. Chen, H. Sun, X. Zhang, and D. Korošak, "Anomalous diffusion modeling by fractal and fractional derivatives," *Computers & Mathematics with Applications*, vol. 59, no. 5, pp. 1754–1758, 2010.
- [31] Y. Xue, S. Rodriguez, and P. Bogdan, "A spatio-temporal fractal model for a CPS approach to brain-machine-body interfaces," in *Proceedings of the 2016 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2016, pp. 642–647.
- [32] O. Romero, S. Chatterjee, and S. Pequito, "Fractional-order model predictive control for neurophysiological cyber-physical systems: A case study using transcranial magnetic stimulation," in *Proceedings of the 2020 American Control Conference (ACC)*. IEEE, 2020, pp. 4996–5001.
- [33] S. Chatterjee, O. Romero, A. Ashourvan, and S. Pequito, "Fractional-order model predictive control as a framework for electrical neurostimulation in epilepsy," *Journal of Neural Engineering*, vol. 17, no. 6, 2020, Art. no. 066017.
- [34] R. L. Magin, *Fractional Calculus in Bioengineering*. Begell House Redding, 2006.
- [35] A. Dzieliński and D. Sierociuk, "Adaptive feedback control of fractional order discrete state-space systems," in *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, vol. 1, Nov 2005, pp. 804–809.
- [36] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, 1992.
- [37] K. Sikorski, "Bisection is optimal," *Numerische Mathematik*, vol. 40, no. 1, pp. 111–117, 1982.
- [38] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
- [39] A. N. Khambhati, K. A. Davis, B. S. Oommen, S. H. Chen, T. H. Lucas, B. Litt, and D. S. Bassett, "Dynamic network drivers of seizure generation, propagation and termination in human neocortical epilepsy," *PLoS Computational Biology*, vol. 11, no. 12, 2015, Art. no. e1004608.
- [40] A. Ashourvan, S. Pequito, A. N. Khambhati, F. Mikhail, S. N. Baldassano, K. A. Davis, T. H. Lucas, J. M. Vettel, B. Litt, G. J. Pappas, et al., "Model-based design for seizure control by stimulation," *Journal of Neural Engineering*, vol. 17, no. 2, 2020, Art. no. 026009.
- [41] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [42] —, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, <http://stanford.edu/~boyd/graph.dcp.html>.
- [43] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.
- [44] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.
- [45] C. Wu, C. Yang, H. Zhao, and J. Zhu, "On the convergence of the EM algorithm: A data-adaptive analysis," *arXiv preprint arXiv:1611.00519*, 2016.
- [46] S. Balakrishnan, M. J. Wainwright, and B. Yu, "Statistical guarantees for the EM algorithm: From population to sample-based analysis," *The Annals of Statistics*, vol. 45, no. 1, pp. 77–120, 2017.
- [47] B. Yan, M. Yin, and P. Sarkar, "Convergence of gradient EM on multi-component mixture of Gaussians," in *Advances in Neural Information Processing Systems*, 2017, pp. 6956–6966.
- [48] J. L. Adams, T. T. Hartley, and C. F. Lorenzo, "Fractional-order system identification using complex order-distributions," *IFAC Proceedings Volumes*, vol. 39, no. 11, pp. 200–205, 2006.
- [49] A. Dzieliński, D. Sierociuk, G. Sarwas, I. Petráš, I. Podlubny, and T. Škovránek, "Identification of the fractional-order systems: A frequency domain approach," *Acta Montanistica Slovaca*, 2011.
- [50] G. Gupta, S. Pequito, and P. Bogdan, "Dealing with unknown unknowns: Identification and selection of minimal sensing for fractional dynamics with unknown inputs," in *Proceedings of the 2018 American Control Conference (ACC)*. IEEE, 2018, pp. 2814–2820.