Selecting Sensors in Biological Fractional-Order Systems

Vasileios Tzoumas, *Student Member, IEEE*, Yuankun Xue, Sérgio Pequito, *Member, IEEE*, Paul Bogdan, *Member, IEEE*, and George J. Pappas, *Fellow, IEEE*

***Abstract*—In this paper, we focus on sensor selection, that is, determine the minimum number of state variables that need to be measured, to monitor the evolution of the entire biological system, that is, all state variables, when modeled by discrete-time fractional-order systems (DTFOS) that are subject to modeling errors, process, and measurement noise. These systems are particularly relevant while modeling the spatiotemporal dynamics of processes in which the impact of long-range memory cannot be properly modeled by multivariate autoregressive integrative moving-average models. Therefore, the DTFOS enable a unified state-space framework to model the evolution of several biological (e.g., stem cell growth and bacteria evolution) and physiological signals (e.g., electroencephalogram and electromyogram). Therefore, in this paper, we focus on the solution to four different (yet related) problems of sensor selection for DTFOS, which are motivated by constraints on the data acquisition that are enforced by the detrimental impact of the sensing mechanisms to the biological system, the cost of performing the measurements with the current sensing technology, or spatial constraints that limit the number of sensors that can be deployed. Toward determining the solution to these problems that we show to be NP-hard, we leverage the representation of the DTFOS to derive new objectives and conditions that, ultimately, enable us to efficiently approximate a solution to the different problems by exploiting the submodularity structure, which enables us to establish suboptimality guarantees.**

***Index Terms*—Biological techniques, biomedical signal processing, electroencephalography, fractional calculus, Kalman filters, observability.**

Manuscript received September 22, 2017; revised January 14, 2018; accepted January 21, 2018. Date of publication February 27, 2018; date of current version June 18, 2018. This work was supported in part by the TerraSwarm Research Center, one of six centers supported by the STARnet phase of the Focus Center Research Program, a Semiconductor Research Corporation Program sponsored by MARCO and Defense Advanced Research Projects Agency (DARPA), in part by AFOSR Complex Networks Program, and in part by AFOSR MURI CHASE. The work of P. Bogdan and Y. Xue was supported in part by the U.S. Army Defense Advanced Research Projects Agency (DARPA) and in part by DARPA Young Faculty Award under Grant W911NF-17-1-0076 and Grant N66001-17-1-4044, and in part by the U.S. National Science Foundation under CAREER Award CPS-1453860. Recommended by Associate Editor Adilson E. Motter. *(Corresponding author: Sergio Pequito.)*

V. Tzoumas and G. J. Pappas are with the Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: vtzoumas@seas.upenn.edu; pappasg@seas.upenn.edu).

S. Pequito is with the Industrial and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12180-3590 USA (e-mail: goncas@rpi.edu).

Y. Xue and P. Bogdan are with the USC Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089-0911 USA (e-mail: yuankunx@usc.edu; paulbogdan2012@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCNS.2018.2809959

I. Introduction

A MULTITUDE of complex systems exhibits long-range (nonlocal) properties, interactions, and/or dependencies (e.g., power-law decays in the weights of linear combination of past data) used to describe the biological system evolution. An example of such systems includes Hamiltonian systems, where the memory (i.e., dependence on the past data) is the result of stickiness of trajectories in time to the islands of regular motion [1]. Alternatively, it has been rigorously confirmed that the viscoelastic properties are typical for a wide variety of biological entities like stem cells, liver, pancreas, heart valve, brain, muscles [1]–[9], suggesting that the long-range memory of these systems obeys the power law distributions. These dynamical systems can be characterized by the well-established mathematical theory of fractional calculus [10], and the corresponding systems could be described by fractional-differential equations [11]–[15]. However, it is until recently that the fractional-order system (FOS) started to find its strong position in a wide spectrum of applications in different domains due to the availability of computing and data-acquisition methods to evaluate its efficacy in terms of capturing the underlying system states evolution.

Specifically, in [11], by the adoption of non-Gaussian statistical approaches, the authors identify the coexistence of fast and slow dividing subpopulations, and quiescent cells, in stem cells from three species. The mathematical analysis also shows that instead of developing independently, stem cells exhibit a time-dependent fractal behavior as they interact with each other through molecular and tactile signals. In [12], the existence of a statistical fractal behavior and inadequacy of modeling blood glucose dynamics via linear state-space models is proved by the multifractal spectrum computed from the blood glucose time series of four individuals. An FOS model is alternatively proposed and evaluated to be superior regarding predictive power and controller synthesis. In [13], a multidimensional FOS is considered to capture the muscular dynamics in the process of forearm movement. The motivation comes from the power-law correlation decay as opposed to the exponential law, which is fundamentally assumed by the popular autoregressive moving average model. After the retrieval of the FOS model from the observations, it is shown that the model output is superior to ARMA to capture the observed spatial and temporal long-range dependence. In [14], a more comprehensive set of physiological processes (i.e., neural, muscular, and vascular processes) are considered to study the minimal sensor placement problem

in the context of the multidimensional FOS. The experimental results suggest that the adoption of FOS and the control theory developed based on it can help improve the design of efficient and reliable cyber-physical systems in the biomedical domain. In [15], Xue and Bogdan propose a statistical nonextensive causal inference framework and construct the generalized master equation (GME) to characterize the dynamics of complex systems that exhibit power-law spatiotemporal dependencies. The solution of the GME suggests a FOS should be considered to capture the dynamical behaviors of the systems. In addition to the application of fractional-order calculus to differentiable dynamical systems, very recent efforts have also been very successful to extend local fractional calculus to nondifferentiable, irregular sets such as fractals or fractal networks [16]–[19]. The fractality/multifractality of network, their characterization, computation, and their influence on the dynamics of complex networked systems is attracting greater attention from a multidisciplinary perspective. The possibility to extend the fractional to self-similar nonsmooth objects is opening new frontiers in science. Nonlinear analysis of data offers still unsolved analytical problems related not only to complex physics and abstract mathematical theories including fractals and fractional calculus [20].

Subsequently, because the current sensing technology is mainly digital, we focus on *discrete-time fractional-order systems* (DTFOS) [21], whose parameterization consists of a relatively small number of parameters, and the dynamics subject to modeling errors and external disturbances. Furthermore, in addition to modeling errors and external disturbances in the DTFOS dynamics, we also account for external disturbances in the sensing technology since our motivating technology, that is, the EEG, uses the sensing technology where noise commonly corrupts the collected data. Subsequently, in this paper, we propose exploring and exploiting the tradeoffs between the selected sensors and the capability to assess the process state over time, which we refer to as estimation performance, since the state is obtained up to a confidence level subject to disturbance and noise. In other words, the combined effect of the modeling errors and external disturbances in the spatiotemporal dynamical processes requires proper deployment of sensing technology that guarantees the best estimation performance (that is, the least estimation error) given the modeling errors' and external disturbances' characteristics.

In the last few years, we have witnessed a growing interest on the tradeoff between the number of used sensors and the degree of observability of linear time-invariant (LTI) systems [22]–[28], which are a particular case of DTFOS. In particular, this tradeoff has been explored under the assumption that either the exact LTI system model is available, in which case one needs to ensure observability [22]–[28]. More recently, this interest is extended to deal with DTFOS, either when the models are exact [14], or in the context of structural observability [29]. Although ensuring that the observability is a key toward the implementation of stable estimators, it does not explicitly explore the tradeoffs between the chosen sensors and the quality of the state estimate and the model uncertainty, which is of utmost importance in the biological settings, for example, in EEG applications. These tradeoffs have been studied so far only for LTI systems, as we briefly review next. In [30] and [31], Zhang *et al.* and Belabbas explore the tradeoffs for LTI systems in the context of Kalman estimation. Specifically, in [30], Zhang *et al.* consider placing small numbers of sensors to optimize the resulting minimum mean square error (MMSE), and in [31], Belabbas designs an output matrix with a desired norm that minimizes the MMSE.

In this paper, we extend the current literature to address the tradeoff between the chosen sensors and the quality of the state estimate for the case of DTFOS with a known parametric model and under possible uncertainties in the dynamics, as well as noise in the measurements collected by sensing technology. Specifically, we address the following problems:

1) determine the minimum number of sensors to ensure bounded process disturbance error within a prescribed threshold;
2) determine the placement of a specified number of sensors to minimize the process disturbance error;
3) determine the minimum number of sensors to ensure bounded state estimation error within a prescribed threshold;
4) determine the placement of a specified number of sensors to minimize the state estimation error.

It is worth noticing that among these four problems, the first couple of problems enforces the validity of the model by quantifying the uncertainty of the system's evolution, whereas the remaining two aim to determine the most likely state of the process across a time window.

The main contributions of this paper can be cast in the following three domains:

Translational: It equips scientists (e.g., biologists and neuroscientists) and engineers alike with a unified framework to decide upon the sensor measurements to be considered to perform state estimation, that is, to perform *sensor selection* to quantify uncertainty in the state and unknown disturbances and noises, and in the context of fractional-order state-space representations capable of modeling spatiotemporal dynamics of processes in which the impact of long-range memory cannot be properly modeled by multivariate autoregressive integrative moving-average models.

Theoretical: We propose deriving observability conditions that enable the quantification of the uncertainty of biological processes modeled by the proposed state-space representation, as well as identifying the state variables that play a key role in monitoring the evolution of the dynamics while making the tradeoff with the accuracy of the estimation. Specifically, we propose computationally efficient algorithms to provide suboptimal solutions to the minimum number of variables that need to be measured (that is NP-hard), while establishing guarantees on the optimality gap.

Application: Recently, there is a renewed interest in neurowearable devices largely boosted by initiatives sponsored by either the Facebook that aims to use wearable devices to write 100 words per minute, and the NeuraLink by Elon Musk that aims to develop implantable brain–computer interfaces. Subsequently, we propose revisiting the neurowearables that rely on electroencephalogram, and determine the sensor location that

seems to be the most effective with respect to a prespecified number of sensors. In particular, we argue that for a variety of tasks, the location of sensors currently used in such wearable devices is suboptimal with respect to the proposed objectives that aim to ensure the quality of the estimated state, process, and measurement noise. Consequently, we conclude that at the light of this framework, some of the neurowearables should be redesigned to enhance the dynamic systems properties such as observability.

In summary, our main contributions are as follows:

1) we formalize the sensor placement problems in the context of four different (yet related) problems pertaining to sensor placement to minimize the process disturbance error and state estimation error;
2) we show that these problems are NP-hard;
3) we present approximation schemes for their solution that have provably optimal approximation performance;
4) we illustrate the proposed approaches using EEG signal data associated with a variety of tasks.

The remainder of this paper is organized as follows. In Section II, we provide our setup and problem formulation. In Section III, we present our main results. In Section IV, we illustrate how the main results can be applied in the context of real EEG signal data. Section V concludes this paper.

II. Problem Statement

In this section, we introduce the problems addressed in this paper. First, we introduce the DTFOS model used in Section II-A, while revisiting some of its properties, and the best linear estimator for it in Section II-B. Then, in Section II-C, we introduce the optimal sensor placement problem for DTFOS, which seeks to determine the minimum collection of sensors that ensure a prespecified estimation performance, or the configuration of a given number of sensors that attain the best process disturbance and estimate quality.

A. DTFOS Model

We consider the linear DTFOS described by

$$\begin{aligned} \Delta x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k, \quad k = 0, 1, \dots \end{aligned} \tag{1}$$

where $x_k = [x_k^1,\ x_k^2,\ \dots,\ x_k^n]^\top \in \mathbb{R}^n$ ($n \in \mathbb{N}$) is the state vector, $y_k \in \mathbb{R}^c$ is the measured output vector, w_k is the process disturbance, v_k is the measurement noise, and x_0 is the initial condition. In addition, $\Delta \equiv \operatorname{diag}(\Delta_{k+1}^{\alpha_1}, \Delta_{k+1}^{\alpha_2}, \dots, \Delta_{k+1}^{\alpha_n})$ is the diagonal matrix operator, where $\Delta_{k+1}^{\alpha_i}$ is the discrete fractional-order difference operator such that

$$\Delta_{k+1}^{\alpha_i} x_{k+1}^i \equiv \sum_{j=0}^{k+1} (-1)^j \binom{\alpha_i}{j} x_{k-j+1}^i$$

and $\binom{\alpha_i}{j} = \frac{\Gamma(\alpha_i+1)}{\Gamma(j+1)\Gamma(\alpha_i-j+1)}$, where $\alpha_i > 0$ is the *fractional-order exponent*, and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$ is the Gamma function. In summary, the matrix A captures the spatial coupling (i.e., dependence) of the process, whereas α_i captures the temporal dependency of the process associated with x_i.

Also, we notice that it is possible to provide a closed-form solution to (1), following [32], and which can be described as follows.

Lemma 1: For all $k \geq 1$, the solution to (1) is given by $x_k = G_k x_0 + \sum_{j=0}^{k-1} G_{k-1-j} w_j$, where

$$G_k \equiv \begin{cases} I, & k = 0 \\ \sum_{j=0}^{k-1} A_j G_{k-1-j}, & k \geq 1 \end{cases}$$

where $A_0 = A$, and A_j is a diagonal matrix whose ith entry is $(-1)^j \binom{\alpha_i}{j+1}$. ◇

In particular, Lemma 1 states that a linear DTFOS can be interpreted as a linear time-variant switching system, where transitions are known. Subsequently, we can develop a Kalman-like estimator for this process, whose estimates' characterization is leveraged to study the tradeoffs between the performance of the estimator and a specified sensor placement.

B. Minimum Variance Linear Estimator

For any estimation horizon K (that is, k in (1) varies from 0 to K), we first present the minimum mean square linear estimator of $z_K \equiv (x_0^\top, w_0^\top, w_1^\top, \dots, w_{K-1}^\top)^\top$. This estimator is particularly useful in biological systems to assess the validity of the model, since a quantification of uncertainty is obtained. To this end, we use the following common assumption.

Assumption 1: Let the initial condition be unknown and modeled by a random variable whose expected value is $\bar{x}_0$ and its covariance is $\mathbb{C}(x_0) \succ 0$. In addition, let the process disturbance w_k and the measurement noise v_k to be described by zero-mean random variables, whose covariance is described, respectively, by $\mathbb{C}(w_k) \succ 0$ and $\mathbb{C}(v_k) \succ 0$, for all $k \geq 0$, where $\mathbb{C}(v_k)$ is a diagonal matrix; that is, the measurement noises between any two sensors that correspond to two rows of C are uncorrelated. Furthermore, for all $k, k' \geq 0$ with $k \neq k'$, let the x_0, w_k, and v_k, as well as the w_k, $w_{k'}$, v_k, and $v_{k'}$ be uncorrelated. ◦

Moreover, we consider the following notations: let the vector of measurements $y_{0:K} \equiv (y_0^\top, y_1^\top, \dots, y_K^\top)^\top$, the vector of process noises $w_{0:K-1} \equiv (w_0^\top, w_1^\top, \dots, w_{K-1}^\top)^\top$, and the vector of measurement noises $v_{0:K} \equiv (v_0^\top, v_1^\top, \dots, v_K^\top)^\top$. Notice that whereas the vector $y_{0:K}$ is known, the vectors $w_{0:K-1}$ and $v_{0:K}$ are not. In addition, we refer to the interval $[0, K] \equiv \{0, 1, \dots, K\}$ as the *estimation horizon* of (1), and its *length* is $K + 1$.

Next, given an estimation horizon $[0, K]$, to derive the minimum mean square linear estimator of z_K, from (1) and Lemma 1, we have

$$y_{0:K} = \mathcal{O}_K z_K + v_{0:K} \tag{2}$$

where $\mathcal{O}_K = [L_0^\top C^\top, L_1^\top C^\top, \dots,\ L_K^\top C^\top]^\top$ with the $n \times n(K+1)$ matrix $L_i = [G_i, G_{i-1}, \dots, G_0, \mathbf{0}]$, and $\mathbf{0}$ is the zero matrix with appropriate dimensions.

Thus, following similar steps to those performed for LTI systems [33], the minimum mean square linear estimate of z_K is

given by

$$\hat{z}_K \equiv \mathbb{E}(z_K) + \mathbb{C}(z_K)\mathcal{O}_K^\top(\mathcal{O}_K\mathbb{C}(z_K)\mathcal{O}_K^\top + \mathbb{C}(v_{0:K}))^{-1} (y_{0:K} - \mathcal{O}_K\mathbb{E}(z_K) - \mathbb{E}(v_{0:K}))$$

where $\mathbb{E}(x)$ is the expected value of x and $\mathbb{C}(x) \equiv \mathbb{E}([x - \mathbb{E}(x)][x - \mathbb{E}(x)]^\top)$ its covariance. Furthermore, the error covariance of $\hat{z}_K$ is given by

$$\begin{aligned}\Sigma_{\hat{z}_K} &\equiv \mathbb{E}((z_K - \hat{z}_K)(z_K - \hat{z}_K)^\top) \\ &= \mathbb{C}(z_K) - \mathbb{C}(z_K)\mathcal{O}_K^\top(\mathcal{O}_K\mathbb{C}(z_K)\mathcal{O}_K^\top + \mathbb{C}(v_{0:K}))^{-1} \mathcal{O}_K\mathbb{C}(z_K).\end{aligned} \tag{3}$$

In this paper, we capture the estimation performance of $\hat{z}_K$ with the metric $\log\det(\Sigma_{\hat{z}_K})$, which is proportional to the conditional entropy of z_K given the measurements $y_{0:K}$ and, as a result, captures how well z_K is explained by $y_{0:K}$ [34, Prop. 2]. In particular, the metric $\log\det(\Sigma_{\hat{z}_K})$ captures the probability that the estimation error $\|z_K - \hat{z}_K\|_2^2$ is small. To explain this, consider the η-confidence ellipsoid of $z_K - \hat{z}_K$ [35]: The η-confidence ellipsoid is the minimum volume ellipsoid that contains $z_K - \hat{z}_K$ with probability η. Specifically, it is encapsulated by $\mathcal{E}_\epsilon(\hat{z}_K) \equiv \{z : z^\top\Sigma_{\hat{z}_K} z \le \epsilon\}$, where $\epsilon \equiv F^{-1}_{\chi^2_{n(K+1)}}(\eta)$ and $F_{\chi^2_{n(K+1)}}$ is the cumulative distribution function of a χ-squared random variable with $n(K+1)$ degrees of freedom [36]. Therefore, the volume of $\mathcal{E}_\epsilon(\hat{z}_K)$ that quantifies the estimation's error of $\hat{z}_K$ is given as follows:

$$\text{vol}(\mathcal{E}_\epsilon(\hat{z}_K)) \equiv \frac{(\epsilon\pi)^{n(K+1)/2}}{\Gamma\left(n(K+1)/2+1\right)} \det\left(\Sigma_{\hat{z}_K}^{1/2}\right). \tag{4}$$

Henceforth, if we consider the logarithm of (4), we obtain

$$\log \text{vol}(\mathcal{E}_\epsilon(\hat{z}_K)) = \beta + 1/2\log\det(\Sigma_{\hat{z}_K}) \tag{5}$$

where β is a constant that depends only on $n(K+1)$ and ϵ, and, as a result, we refer to the $\log\det(\Sigma_{\hat{z}_K})$ as the $\log\det$ *initial state-uncertainty estimation error of the minimum variance linear estimator* of (1).

Alternatively, we might be interested to determine the minimum variance linear estimator of $x_{0:K} \equiv (x_0, x_1, \ldots, x_K)$ denoted by $\hat{x}_{0:K}$. To this end, the collection of measurements is given by $y_{0:K} = O_K x_{0:K} + v_{0:K}$, where O_K is the block-diagonal matrix with diagonal elements $K+1$ copies of the matrix C. Subsequently, following similar steps to those in [37], the state estimation $\hat{x}_{0:K}$ error covariance is given by

$$\Sigma_{\hat{x}_{0:K}} = \mathbb{C}(x_{0:K}) - \mathbb{C}(x_{0:K})O_K^\top(O_K\mathbb{C}(x_{0:K})O_K^\top + \mathbb{C}(v_{0:K}))^{-1}O_K\mathbb{C}(x_{0:K}). \tag{6}$$

Besides, by proceeding similarly to the reasoning above, we can define the $\log\det$ *batch-state estimation error of the minimum variance linear estimator* of (1) as follows:

$$\log \text{vol}(\mathcal{E}_\epsilon(\hat{x}_{0:K})) = \beta + 1/2\log\det\left(\Sigma_{\hat{x}_{0:K}}\right). \tag{7}$$

C. Optimal Sensor Placement

Now, we introduce four different (yet, related) problems to assess the optimal sensor placement with respect to the $\log\det$ of the initial state-uncertainty and batch-state estimation error of the minimum variance linear estimator of (1). Specifically, we propose for each to determine the placement of r sensors such that the overall estimation error is minimized, and determine a placement of sensors such that the estimation error satisfies a specified threshold.

Therefore, we propose using the following sensor placement model: across the estimation horizon $[0, K]$, a *unique* subset of r sensors in (1) is placed and used, that corresponds to r of the c rows of C ($r \le c$). In particular, for all $k \in [0, K]$ in (1)

$$y_k = SCx_k + v_k,\ k \in [0, K] \tag{8}$$

where S is the sensor placement matrix (constant across the estimation horizon $[0, K]$); that is, it is a zero-one matrix such that $S_{ij} = 1$ if sensor j is placed (which corresponds to the jth row of C), and $S_{ij} = 0$ otherwise. We assume that a sensor can be placed at most once and, as a result, for each i there is one j such that $S_{ij} = 1$ while for each j, there is at most one i such that $S_{ij} = 1$. Hence, given a sensor selection matrix S, the indices of the rows of C that correspond to used sensors are denoted by $\mathcal{S}$, that is, $\mathcal{S} \equiv \{j :$ exists i such that $S_{ij} = 1\}$.

Consequently, given the DTFOS in (1) and a finite estimation horizon $[0, K]$, we consider the following four problems.

Initial State-Uncertainty Estimation Error 1) Provided a specified error threshold $R \in \mathbb{R}^+$, determine the *initial state-uncertainty minimal sensor placement problem* that is a solution to the following problem:

$$\begin{aligned}\underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{minimize}} \quad & |\mathcal{S}| \\ \text{subject to} \quad & \log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right) \le R\end{aligned} \tag{$\mathcal{P}_1$}$$

where $\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right)$ is the determinant of $\Sigma_{\hat{z}_K}$ in (3) when $\mathcal{O}_K$ is replaced by $\mathcal{O}_K(\mathcal{S})$, with explicit dependence on $\mathcal{S}$, and described by $\mathcal{O}_K(\mathcal{S}) = [L_0^\top C(\mathcal{S})^\top, L_1^\top C(\mathcal{S})^\top, \ldots, L_K^\top C(\mathcal{S})^\top]^\top$, and $C(\mathcal{S})$ denotes the rows of C with indices in $\mathcal{S}$; 2) Provided a maximum number r of sensors to be placed, determine the *initial state-uncertainty cardinality-constrained sensor placement problem for minimum estimation error* that consists of a solution to the following problem:

$$\begin{aligned}\underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{minimize}} \quad & \log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right) \\ \text{subject to} \quad & |\mathcal{S}| \le r.\end{aligned} \tag{$\mathcal{P}_2$}$$

Batch-State Estimation Error 3) Provided a specified error threshold $R \in \mathbb{R}^+$, determine the *minimal sensor placement problem* that is a solution to the following problem:

$$\begin{aligned}\underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{minimize}} \quad & |\mathcal{S}| \\ \text{subject to} \quad & \log\det\left(\Sigma_{\hat{x}_{0:K}}(\mathcal{S})\right) \le R\end{aligned} \tag{$\mathcal{P}_3$}$$

where $\det\left(\Sigma_{\hat{x}_{0:K}}(\mathcal{S})\right)$ is the determinant of $\Sigma_{\hat{x}_{0:K}}$ in (6) when O_K is replaced by $O_K(\mathcal{S})$, which is the block-diagonal matrix with diagonal elements K copies of the matrix $C(\mathcal{S})$, and $C(\mathcal{S})$ denotes the rows of C with indices in $\mathcal{S}$; and 4) Provided a

maximum number r of sensors to be placed, determine the *cardinality-constrained sensor placement problem for minimum estimation error* that consists of a solution to the following problem:

$$\begin{array}{ll}\underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{minimize}} & \log\det\left(\Sigma_{\hat{x}_{0:K}}(\mathcal{S})\right)\\ \text{subject to} & |\mathcal{S}|\leq r.\end{array} \qquad (\mathcal{P}_4)$$

$\circ$

Problems $(\mathcal{P}_1)-(\mathcal{P}_4)$ address different problems that focus on different practical considerations. Specifically, $(\mathcal{P}_1)$ aims to determine the minimum number of sensors to ensure bounded process disturbance error within a prescribed threshold, which enables the minimization of the estimation of the uncertainty that drives the system; thus, equipping us with an uncertainty quantification of the process evolution. In contrast, $(\mathcal{P}_2)$ addresses the problem of determining the placement of a specified number of sensors to minimize the process disturbance error, which captures the situations where one has a budget on the available sensing technology, and wants to deploy the sensors to maximize the performance of the process captured by minimizing the system's uncertainty.

Problem $(\mathcal{P}_3)$ focuses on determining the minimum number of sensors to ensure bounded state estimation error within a prescribed threshold, which might be related to the satisfaction of some standard or accuracy required to have a sound estimate of the system's state. Finally, $(\mathcal{P}_4)$ targets the placement of a specified number of sensors to minimize the state estimation error when the number of sensing mechanisms is limited and one aims to minimize the system's state estimate uncertainty.

Notwithstanding, as it will become clear in the upcoming sections, the underlying optimization structure is similar, which enables us to study them in a unified fashion. Specifically, to address these problems, we will show that both $\log\det(\Sigma_{\hat{z}_K})$ and $\log\det(\Sigma_{\hat{x}_{0:K}})$ are *supermodular* and *nonincreasing* (formally defined in Section III). As a consequence, approximation algorithms for these types of functions can be leveraged to provide approximate solutions to these problems with worst-case performance guarantees.

III. Sensor Placement for DTFOS

We present the main results of this paper. First, we show that $(\mathcal{P}_1)$–$(\mathcal{P}_4)$ are NP-hard (see Theorem 1), which implies that optimal polynomial solutions to these problems are unlikely to exist. Next, we propose polynomial algorithms (see Algorithms 1 and 2) to obtain an approximate solution to these problems, while ensuring worst-case performance guarantees (see Theorem 3 and 4). In more detail, in Theorem 2, we show that the constraint and objective function in $(\mathcal{P}_1)/(\mathcal{P}_3)$ and $(\mathcal{P}_2)/(\mathcal{P}_4)$, respectively, are supermodular. Thereby, greedy algorithms can be provided to approximate the solution to these problems while ensuring a worst-case scenario bounded optimality gap. Finally, in Theorem 5, we provide a discussion on the fundamental limits on the state-uncertainty estimation error and batch-state estimation, while exploring the tradeoff with problems' parameters.

We start by showing the computational complexity of our problems in the next result.

Theorem 1: The problems $(\mathcal{P}_1)$–$(\mathcal{P}_4)$ are NP-hard. $\diamond$

Subsequently, we need to devise a strategy that approximates the solutions to the proposed problems. Toward this goal, consider the following definitions.

Definition 1: A function $h: 2^{[c]}\mapsto\mathbb{R}$ is *submodular*, where $[c]=\{1,\ldots,c\}$, if for any sets $\mathcal{S}$ and $\mathcal{S}'$, with $\mathcal{S}\subseteq\mathcal{S}'\subseteq[c]$, and any $a\notin\mathcal{S}'$,

$$h(\mathcal{S}\cup\{a\})-h(\mathcal{S})\geq h(\mathcal{S}'\cup\{a\})-h(\mathcal{S}').$$

A function $h: 2^{[c]}\mapsto\mathbb{R}$ is *supermodular*, if $(-h)$ is submodular. $\diamond$

Definition 2: A function $h: 2^{[c]}\mapsto\mathbb{R}$ is a *nonincreasing set function*, if for any $\mathcal{S}\subseteq\mathcal{S}'\subseteq[c]$ it follows that $h(\mathcal{S})\geq h(\mathcal{S}')$. Moreover, h is a *nondecreasing set function* if $(-h)$ is a nonincreasing set function. $\diamond$

Furthermore, a function $h: 2^{[c]}\mapsto\mathbb{R}$ is submodular if, for any $a\in[c]$, the function $h_a: 2^{[c]\setminus\{a\}}\mapsto\mathbb{R}$ defined as $h_a(\mathcal{S})\equiv h(\mathcal{S}\cup\{a\})-h(\mathcal{S})$ is a nonincreasing set function. This property is commonly referred to as the *diminishing returns property* [38].

Now, we show that the constraint and objective function of $(\mathcal{P}_1)/(\mathcal{P}_3)$ and $(\mathcal{P}_2)/(\mathcal{P}_4)$, respectively, are supermodular and nonincreasing.

Theorem 2: Let c be the number of rows of C, and $s_i\in\{0,1\}$ be 1 if and only if ith sensor (ith row of C) is placed, and $L_{0:K}\equiv[L_0^\top,L_1^\top,\ldots,L_K^\top]^\top$. In addition, let $M^{(i)}\equiv C_{0:K}^\top I^{(i)}\mathbb{C}(v_{0:K})^{-1}I^{(i)}C_{0:K}$, where $C_{0:K}$ is the block diagonal matrix where each of its $K+1$ diagonal elements is equal to C, and $I^{(i)}$ is the diagonal matrix with $c(K+1)$ diagonal elements such that, for all $k\in[0,K]$, the $(kc+i)$th element is 1, and the rest of the elements are equal to zero. Then, given any finite estimation horizon $[0,K]$, the following two equalities hold:

$$\log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right)=-\log\det\left(\sum_{i=1}^{c}s_iL_{0:K}^\top M^{(i)}L_{0:K}+\mathbb{C}(z_K)^{-1}\right)$$

and

$$\log\det\left(\Sigma_{\hat{x}_{0:K}}(\mathcal{S})\right)=-\log\det\left(\sum_{i=1}^{c}s_iM^{(i)}+\mathbb{C}(x_{0:K})^{-1}\right).$$

Furthermore, both $\log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right)$ and $\log\det\left(\Sigma_{\hat{x}_{0:K}}(\mathcal{S})\right)$ are supermodular and nonincreasing set functions with respect to the choice of the sensor set $\mathcal{S}\subset[c]=\{1,\ldots,c\}$. $\diamond$

As a consequence of Theorem 2, it follows that the functions exhibit the diminishing returns property, i.e., its rate of reduction with respect to newly placed sensors decreases as the cardinality of the already placed sensors increases. Therefore, some well-known approximation schemes [39], [40] can be leveraged to obtain suboptimal solutions to $(\mathcal{P}_1)$–$(\mathcal{P}_4)$ with optimality guarantees.

In Algorithms 1 and 2, we present strategies to approximate the solutions to $(\mathcal{P}_1)/(\mathcal{P}_3)$ and $(\mathcal{P}_2)/(\mathcal{P}_4)$, respectively. Specifically, in Algorithm 1, we provide an efficient algorithm for $(\mathcal{P}_1)/(\mathcal{P}_3)$ that returns a sensor set that satisfies the prescribed threshold and has cardinality up to a multiplicative factor from

Algorithm 1: Approximation Algorithm for $(\mathcal{P}_1)$/$(\mathcal{P}_3)$.

Input: $h_\alpha(\mathcal{S}) = \log\det(\Sigma_\alpha(\mathcal{S}))$, where $\alpha \in \{\hat{x}_{0:K}, \hat{z}_K\}$ for $k \in [0, K]$, and a threshold R on the total estimation error incurred by $h_\alpha(\mathcal{S})$.
Output: Approximate solution $\mathcal{S}_\alpha$ for $(\mathcal{P}_1)$/$(\mathcal{P}_3)$.
$\mathcal{S}_\alpha \leftarrow \emptyset$
while $h_\alpha(\mathcal{S}_\alpha) > R$ **do**
$a_i \leftarrow a' \in \arg\max_{a \in [c]\setminus\mathcal{S}_\alpha} (h_\alpha(\mathcal{S}_\alpha) - h_\alpha(\mathcal{S}_\alpha \cup \{a\}))$
$\mathcal{S}_\alpha \leftarrow \mathcal{S}_\alpha \cup \{a_i\}$
end while

Algorithm 2: Approximation Algorithm for $(\mathcal{P}_2)$/$(\mathcal{P}_4)$.

Input: $h_\alpha(\mathcal{S}) = \log\det(\Sigma_\alpha(\mathcal{S}))$, where $\alpha \in \{\hat{x}_{0:K}, \hat{z}_K\}$ for $k \in [0, K]$, and a bound on the number r of sensors used to minimize $h_\alpha(\mathcal{S})$.
Output: Approximate solution $\mathcal{S}_\alpha$ for $(\mathcal{P}_2)$/$(\mathcal{P}_4)$.
$\mathcal{S}_\alpha \leftarrow \emptyset$, $i \leftarrow 0$
while $i < r$ **do**
$a_i \leftarrow a' \in \arg\max_{a \in [c]\setminus\mathcal{S}_\alpha} (h_\alpha(\mathcal{S}_\alpha) - h_\alpha(\mathcal{S}_\alpha \cup \{a\}))$
$\mathcal{S}_\alpha \leftarrow \mathcal{S}_\alpha \cup \{a_i\}$, $i \leftarrow i + 1$
end while

the minimum cardinality sensor sets that meet the same estimation bound. More importantly, this multiplicative factor depends only logarithmically on the problems' parameters. These properties and the time complexity are described in the following result.

Theorem 3: Let a solution to $(\mathcal{P}_1)$/$(\mathcal{P}_3)$ be denoted by $\mathcal{S}_\alpha^\star$, and the set obtained by Algorithm 1 be denoted by $\mathcal{S}_\alpha$. Moreover, denote the maximum diagonal element of $\mathbb{C}(x_0)$ and $\mathbb{C}(w_k)$, among all $k \in [0, K]$, as σ_0^2 and σ_w^2, respectively. Then,

$$\log\det(\Sigma_\alpha(\mathcal{S}_\alpha)) \leq R \tag{9}$$

and the optimality gap bounded as follows:

$$\frac{|\mathcal{S}_\alpha|}{|\mathcal{S}_\alpha^\star|} \leq 1 + \log\left\{\frac{\log\det(\Sigma_\alpha(\emptyset)) - \log\det(\Sigma_\alpha([c]))}{R - \log\det(\Sigma_\alpha([c]))}\right\} \equiv \eta \tag{10}$$

where $\log\det(\Sigma_{\hat{z}_K}(\emptyset)) \leq n(K+1)\log\max(\sigma_0^2, \sigma_w^2)$.

Furthermore, the time complexity of Algorithm 1 is $O(c^2(nK)^{2.4})$. ◇

Therefore, Algorithm 1 returns a sensor set that meets the estimation threshold of $(\mathcal{P}_1)$/$(\mathcal{P}_3)$. Moreover, the cardinality of this set is up to a multiplicative factor of η from the minimum cardinality sensor sets that meet the same estimation bound. In other words, η is a worst-case approximation guarantee for Algorithm 1. Besides, η depends only logarithmically on the problems' parameters. Additionally, the dependence of η on n, R, and $\max(\sigma_0^2, \sigma_w^2)$ is expected from a design perspective. Specifically, by increasing the state space size n, requesting a better estimation guarantee by decreasing R, or incurring a noise of greater variance, should all push the cardinality of the selected sensor set upward.

Next, in Algorithm 2, we provide an efficient algorithm for $(\mathcal{P}_2)$/$(\mathcal{P}_4)$ that returns a sensor set of cardinality r, where r is chosen by the designer. In the next result, we provide optimality guarantees of the solution obtained with Algorithm 2, as well as the computational complexity incurred by the algorithm.

Theorem 4: Let a solution to $(\mathcal{P}_2)$/$(\mathcal{P}_4)$ be denoted by $\mathcal{S}_\alpha^\star$, and the set obtained by Algorithm 2 be denoted by $\mathcal{S}_\alpha$. Then,

$$\frac{\log\det(\Sigma_\alpha(\mathcal{S}_\alpha)) - \log\det(\Sigma_\alpha(\emptyset))}{\log\det(\Sigma_\alpha(\mathcal{S}_\alpha^\star)) - \log\det(\Sigma_\alpha(\emptyset))} \geq 1 - \frac{1}{e} \tag{11}$$

where the approximation factor $1 - 1/e$ in (11) is the best one can achieve in polynomial time for this problem.

Furthermore, the time complexity of Algorithm 2 is $O(cr(nK)^{2.4})$. ◇

Notice that from Theorem 4 it follows that the approximation quality depends on n, r, and $\max(\sigma_0^2, \sigma_w^2)$ as expected from a design perspective. Specifically, by increasing the state space size n, requesting a smaller sensor set by decreasing r, or incurring a noise of greater variance should all push the quality of the approximation level downward.

A. Limits on Selecting Sensors in DTFOS

Next, we provide explicit bounds on the variance of the state-uncertainty estimation error, while exploring the tradeoff with the following quantities: 1) the length of the estimation horizon $[0, K]$; 2) the number of placed sensors r; and 3) the characteristics of the noises w_k and v_k. In particular, the next result imposes limitations on the assessment of the results that cannot be overcome.

Theorem 5: Let $\sigma_0^{(-1)} \equiv \max_{i\in[n]}[\mathbb{C}(x_0)^{-1}]_{ii}$, $\sigma_w^{(-1)} \equiv \max_{i\in[n(K+1)]}[\mathbb{C}(w_{0:K})^{-1}]_{ii}$, and $\sigma_v^{(-1)} \equiv \|\mathbb{C}(v_{1:K})^{-1}\|_2$. Also, denote by $\bar{L}$ the matrix $L_{0:K}L_{0:K}^\top$. Then, the following inequality holds for the variance of the error of the minimum variance estimator $\hat{z}_K$:

$$\mathrm{tr}(\Sigma_{\hat{z}_K}) \geq \frac{n(K+1)}{r\sigma_v^{(-1)}\|C\|_2^2\|\bar{L}\|_2 + \max\{\sigma_0^{(-1)}, \sigma_w^{(-1)}\}}. \tag{12}$$

◇

In other words, for constant $\|C\|_2^2$ and $\|\bar{L}\|_2$, (12) implies that the state-uncertainty estimation error used to assess the validity of the model (1) is bounded by a quantity that decreases as the number of placed sensors r increases, and increases as the system's state size or the horizon K increases. Subsequently, it implies that $\mathrm{tr}(\Sigma_{\hat{z}_K})$ can decrease only when inversely proportional with the number r of placed sensors, and, as a result, increasing the number r to reduce the variance of the error of the minimum variance linear estimator is ineffective. Additionally, the bound in (12) increases linearly with the system's state size, which imposes additional fundamental limitations for large-scale DTFOS.

Finally, we notice that similar arguments and fundamental bounds can be readily derived for the variance of the batch-state estimator error, i.e., $\mathrm{tr}(\Sigma_{\hat{x}_{0:K}})$ by following the same steps as in [33].

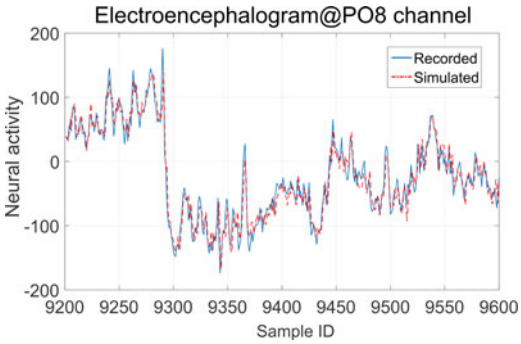

Fig. 1. EEG data recorded and the simulated using DTFOS at the EEG channel PO_8.

IV. EEG Sensor Placement

In this section, we propose to study $(\mathcal{P}_1)$–$(\mathcal{P}_4)$ in a real-world application setting collected by the BCI2000 system with a sampling rate of 160 Hz [41]. Specifically, we consider 64-channel EEG dataset which records the brain activity of ten subjects (S001–S010) when they are performing motor and imagery tasks [42]. Each subject sits in front of a screen where targets might appear at the right/left/top/bottom side of the screen. Upon noticing the target, each subject is asked to open and close the corresponding fists or feet as a function of where the target appears. Each individual performed 14 experimental runs consisting of 1 min with eyes open, 1 min with eyes closed, and three 2 min runs of 4 interacting tasks with the target: (*Task 1*) open and close left or right fist as the target appears on either left or right side of the screen; (*Task 2*) imagine opening and closing left or right fist as the target appears on either left or right side of the screen; (*Task 3*) open and close both fists or both feet as the target appears on either the top or the bottom of the screen; and (*Task 4*) imagine opening and closing both fists or both feet as the target appears on either the top or the bottom of the screen.

First, we estimated the parameters of the DTFOS for the different tasks,[1] which can be modeled by DTFOS as argued in [14]. To illustrate the modeling capabilities of the proposed DTFOS model, in Fig. 1 we contrast the recorded data at location PO_8 against the one simulated using the DTFOS identified. It is worth mentioning that similar performances are achieved across different channels, subjects, and tasks. Besides, the fractional-order exponents range from 0.34 to 1.04 across different tasks, which provides evidence that these could not be properly modeled by LTI systems—see [14] for further details. Finally, we considered that the initial state, disturbance, and measurement noise follow a normal distribution with zero mean and covariance described by the identity matrix (both with appropriate dimensions).

[1]The identification techniques used were introduced in [13], and the software implementation can be found at https://github.com/urashima9616/DFOS_Sensor_Selection.

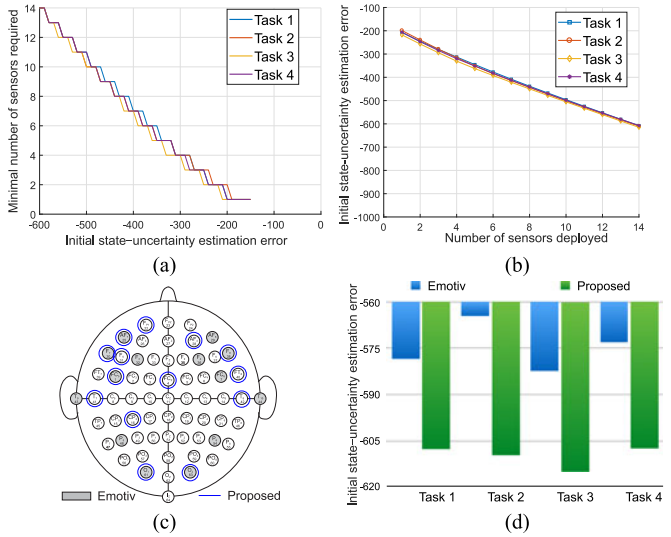

Fig. 2. (a) Minimal sensor placement to achieve a prescribed initial state-uncertainty estimation errors. (b) Initial state-uncertainty $\log\det$ errors achieved given different sensor budgets. (c) Sixty-four-channel geodesic sensor distribution for measurement of EEG, where the sensors in gray represent those of the Emotiv EPOC and the ones in red are those returned by Algorithm 2 when solving $(\mathcal{P}_2)$ (that relieved to be the same for all four tasks), given the identified DTFOS and a deployment budget of 14 sensors. (d) Initial state-uncertainty $\log\det$ estimation errors associated with the highlighted sensor placements in (c).

A. Initial State-Uncertainty Estimation Error

First, we considered a single subject (S002), and determined the different DTFOS systems associated with the four different tasks. We applied Algorithm 1, with $\alpha = \hat{z}_K$ and $K = 7$, to solve $(\mathcal{P}_1)$, and, in Fig. 2(a), we plot the minimal number of sensors required as a function of required initial state-uncertainty $\log\det$ estimation error for the different tasks. The following observations are due: 1) given the same level of initial state-uncertainty estimation error required, the minimal number of sensors varies slightly when the subject is performing different tasks; and 2) given a task, the initial state-uncertainty $\log\det$ error exhibits supermodular properties (see Theorem 2).

To address $(\mathcal{P}_2)$, i.e., to evaluate the achievable levels of the initial state-uncertainty estimation error given different sensor deployment budgets, we resorted to Algorithm 2 with $\alpha = \hat{z}_K$ and $K = 7$. In Fig. 2(b), we present the summary of the results, namely, the $\log\det$ errors given different cardinality-constrains under the four tasks. It can be observed that the information gain, i.e., the improvement on estimation errors, is diminishing as the number of sensors used increases—as predicted by Theorem 4.

Additionally, we considered the deployment of 14 sensors, which is the same number of sensors available in some of the current EEG wearable technology, e.g., the Emotiv EPOC [43]. In Fig. 2(c), we report the sensor deployment returned by Algorithm 2 when solving $(\mathcal{P}_2)$, which revealed to be the same across all four tasks (for the same individual). Specifically, we circled in blue the 14 sensors determined by our framework, whereas the Emotiv EPOC [43] sensors are colored in gray. From Fig. 2(c), we first notice that the sensor distribution pattern of Emotiv EPOC is symmetrical, whereas the Algorithm 2

places the sensors asymmetrically. Moreover, even though some of the locations are fairly close to each other (e.g., 41/43, 16/48, 40/38, 42/44, and 14/7), it turns out that only 5 out of 14 locations (i.e., 25, 29, 44, 48, 63) are identical, and Emotiv EPOC does not consider sensors 23, 4, and 18.

Subsequently, we assessed how the different sensor distributions, i.e., the proposed by our framework and the proposed by Emotiv EPOC, affect the estimation errors. In Fig. 2(d), we report the initial state-uncertainty estimation error across the different tasks. It is worth noticing that the sensors considered by our framework perform considerably better than the sensor distribution used by the Emotiv EPOC. Specifically, the $\log\det$ estimation errors attained by the proposed sensor placement are smaller compared to those of Emotiv EPOC. In fact, it presents considerable gains across the different tasks, and, in particular, in Tasks 2 and 4 that require the use of imagination instead of the motor skills. Finally, it is important to notice that the same sensor placement performs almost equally well across different tasks. Therefore, these results support the fact that from the point-of-view of an initial state-uncertainty estimation error using a model-based approach, the sensors' locations of the commercial EEG devices should be redesigned so to ensure better estimation performance.

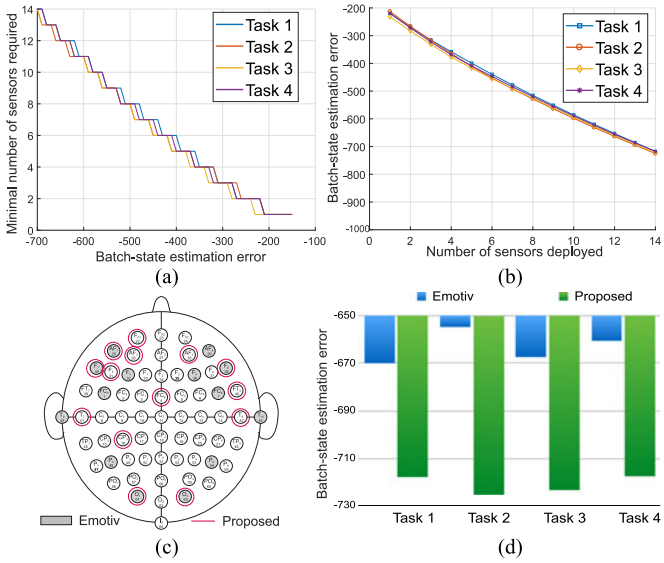

Fig. 3. (a) Minimal sensor placement to achieve a prescribed batch-state estimation errors. (b) Batch-state $\log\det$ errors achieved given different sensor budgets. (c) Sixty-four-channel geodesic sensor distribution for measurement of EEG, where the sensors in gray represent those of the Emotiv EPOC and the ones in red are those returned by Algorithm 2 when solving $(\mathcal{P}_4)$ (that relieved to be the same for all 4 tasks), given the identified DTFOS and a deployment budget of 14 sensors. (d) Batch-state $\log\det$ estimation errors associated with the highlighted sensor placements in (c).

B. Batch-State Estimation Error

Now, we address the batch-state estimation problems proposed in $(\mathcal{P}_3)$–$(\mathcal{P}_4)$. Toward this goal, we consider the DTFOS corresponding to the four different tasks for the same subject (S002). In particular, we obtain the solution to $(\mathcal{P}_3)$ by relying on Algorithm 1 with $\alpha = \hat{x}_{0:K}$ and $K = 7$, whose solutions are found in Fig. 3(a) for several levels of batch-state estimation errors. From Fig. 3(a), we observe that for a specific level of batch-state estimation error, the variation (across the different tasks) in the minimum number of sensors is minor. Moreover, we observe that the batch-state $\log\det$ estimation error exhibits a diminishing returns property—as per Theorem 2.

Next, we use Algorithm 2 with $\alpha = \hat{x}_{0:K}$ and $K = 7$ to tackle problem $(\mathcal{P}_4)$, i.e., to compute the achievable levels of the batch-state estimation error across several sensor placement budgets. The results are presented in Fig. 3(b), where we report the batch-state $\log\det$ estimation errors given different cardinality-constrains across the four tasks. Similarly to the previous figure, we notice that the gain, i.e., the improvement on the estimation error, is diminishing as the number of sensors used increases (see Theorem 4).

Furthermore, using Algorithm 2 to solve $(\mathcal{P}_4)$ with a budget of 14 sensors, we obtained the sensor placement illustrated in Fig. 3(c). In Fig. 3(c), we circle in red the 14 sensors' placement found by our framework, whereas we depict the Emotiv EPOC [43] sensors in gray. Notice that the sensor placement obtained turned out to be the same across all four tasks for the same individual and it is asymmetrical, in contrast with the one of the Emotiv EPOC. Notably, only 5 out of 14 locations (i.e., 25, 29, 44, 48, 63) in both cases are the same, even though some of the sensor locations are close to each other (e.g., 41/43, 16/48, 40/38, 42/44, and 14/7). Moreover, the Emotiv EPOC does not consider the sensors 23, 4, and 18. In Fig. 3(d), we compared the batch-state estimation errors (across the different tasks) of the sensor deployment returned by our framework against the sensors placement of the Emotiv EPOC. We make the following observations: first, the sensors considered by our framework perform considerably better than the sensor distribution used by the Emotiv EPOC, especially with respect to Tasks 2 and 4. Furthermore, it is worth to mention that the same sensor placement returned by our framework performs almost equally well across the different tasks.

C. Assessment of Intersubject Variability

To assess how the intersubject variability of brain dynamics affects the sensor selection under a fixed budget, we next consider a set of experiments where we solve $\mathcal{P}_2$ and $\mathcal{P}_4$ for ten subjects across the four different tasks (Tasks 1–4). In particular, based on the identified DTFOS associated with the four tasks for a subject, we apply Algorithm 2 to obtain the placement of sensors given a deployment budget of 14 sensors by minimizing: 1) initial state-uncertainty estimation error, and 2) batch-state estimation error, respectively. In addition, we also identify the most voted 14 sensor locations based on the poll of suboptimal solutions returned when solving $\mathcal{P}_2$ and $\mathcal{P}_4$ individually for all ten subjects. The results are summarized in Fig. 4(a)–(d). We use heat maps to show the distribution of sensor locations returned by solving $\mathcal{P}_2$ and $\mathcal{P}_4$ individually in Fig. 4(a) and (c), respectively. We can make the following remarks: 1) there exists a noticeable degree of intersubject variabilities in the sensor deployment. This can be evidenced by the fact that the chosen

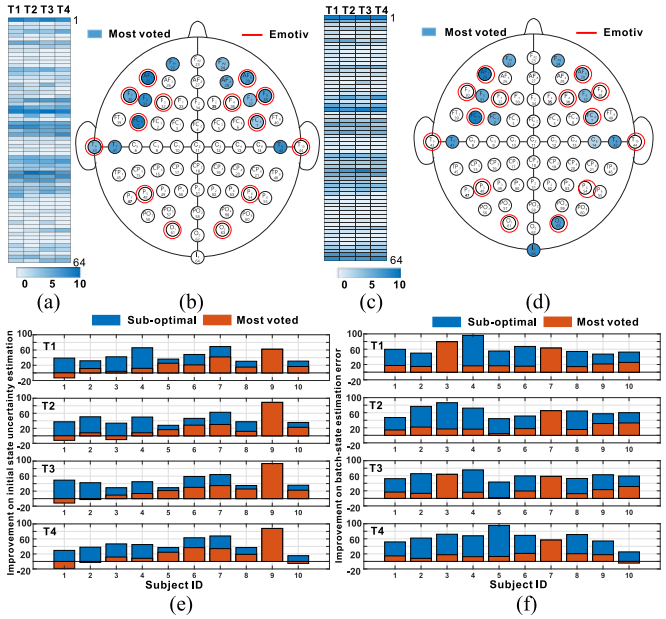

Fig. 4. Sixty-four-channel geodesic sensor distribution over ten subjects under Tasks 1–4 and the most voted deployment given a 14-sensor budget by minimizing (a) the initial state-uncertainty estimation error, and (b) batch-state estimation error. Improvement on (c) initial state-uncertainty estimation error, and (d) batch-state estimation error when: 1) the suboptimal 14-sensor deployment returned by Algorithm 2 individually (blue bar) and, 2) the most voted 14-sensor deployment by ten subjects (red bar) are considered.

sensors span over $40, 41, 46, 46$, and $37, 43, 43, 46$ different locations across the four different tasks when solving $\mathcal{P}_2$ and $\mathcal{P}4$, respectively. This suggests that the underlying brain dynamics are subject to remarkable individual heterogeneities even in response to the same set of tasks. Subsequently, the best-possible sensor schemes have to be designed to be individual-specific.

At the same time, there is a good percentage of agreement and lower loss of performance when planning homogeneous commercial solutions. Specifically, sensors $1, 25, 29, 41, 42$ are almost unanimously chosen by all ten subjects, as a result of solving both $\mathcal{P}_2$ and $\mathcal{P}_4$. This strongly hinges that (for the proposed tasks) there seems to be some fundamental underlying dynamics that enables the state estimation and this subset of sensors are responsible for accessing them. To see this more clearly, we report the top voted 14 sensor locations as an average case and color them in descending order of consensus as a function of darkness in Fig. 4(b) and (d)—as comparison, we also report the sensor deployment proposed by Emotive EPOC with red circles in both geodesic maps. The following observations are due. First of all, notice that all 14 sensors are voted by at least five subjects while sensor $1, 25, 29, 41, 42$ are chosen by at least eight subjects. Second, when considering the minimization of batch-state estimation errors instead of initial state-uncertainty estimation errors, the sensor deployment can be very different. For instance, sensors 63 and 64 are critically important when solving $\mathcal{P}_4$ (as 8 out of 10 subjects choose them) whereas their influence to the initial state-uncertainty estimation are out-weighted by other sensors. Third, similar with our previous case study, only 7/14 and 6/14 most voted sensor deployment by 10 subjects are identical to those proposed by Emotiv EPOC when solving $\mathcal{P}_2$ and $\mathcal{P}_4$, respectively. This potentially suggests the need of the redesign of Emotiv provided the estimation performance we setup in our study, since under our proposed approach the most voted deployment and suboptimal individual deployment (as returned by the algorithms proposed in this paper) achieve better performance. Specifically, we show the improvement on the $\log\det$ estimation error of both initial state-uncertainty and batch state over the one by Emotiv EPOC when the most voted deployment (red bar) and the suboptimal individual deployment (blue bar) are employed—see Fig. 4(e) and (f), respectively. The positive improvement suggests that our proposed deployment is better than that of Emotiv EPOC. We notice that the suboptimal deployment returned by solving $\mathcal{P}_2$ and $\mathcal{P}_4$ individually improves the estimation error significantly in all cases, which is aligned with the results of our previous case study on a single subject. Overall, based on the aforementioned observations one could conclude that the sensors' locations of the commercial EEG devices could be redesigned to enhance both their initial state and batch-state estimation error performance.

D. Discussion of the Results

We first notice that the proposed sensor locations seem to cope better with scenarios where the neuroactivation is not as well understood as the motor-related tasks (Tasks 1 and 3), e.g., imagining actions associated with Tasks 2 and 4. Additionally, we emphasize that the time-window considered was small ($K = 7$), since aimed to attain real-time estimation. As a result, it is expected to obtain even better performance results if we increase the size window. Besides, the cumulative error increases faster in the Emotiv EPOC.

Although in our case study we relied on EEG data, it is expected that the proposed problems have distinct value in different biological settings—as explained in Section I. Specifically, in problems where signals are well modeled by the proposed DTFOS, as it is the case of other physiological signals such as electromyograms (EMG) and electrocardiograms [14].

V. Concluding Remarks

We considered biologically motivated discrete-time linear FOSs and studied the tradeoff between the sensor placement and the properties that pertain to Kalman-like filter performance. Specifically, we formalized the sensor placement problems in context of four different (but related) problems pertaining to the sensor placement to minimize the state-uncertainty and batch-state estimation error. We showed that these problems are NP-hard, and we presented polynomial approximation strategies for their solution that have suboptimality guarantees.

Additionally, we explored the different problems in the context of real EEG data during a period of time where the individuals performed four different tasks. The results obtained support the capability of the proposed framework to deal with critical sensing deployment problems, and unveiled that the number and location of the sensors vary across tasks and subjects for the same experimental setup. Furthermore, we argue that these locations are not compatible with those used by state-of-the-art EEG wearables (e.g., Emotiv EPOC), which supports the need

for further research and redesign of future EEG wearables that aim to attain a specified estimation performance for a given task.

Future research will consider the multiscenario case, where the sensor placement has to consistently and reliably consider possible dynamics, e.g., multitasks simultaneously when EEG is considered. Additionally, we propose to validate the presented methodology when a large cohort of individuals and tasks is considered.

APPENDIX
PROOFS OF THE MAIN RESULTS

Proof of Theorem 1: We prove that $(\mathcal{P}_1)$–$(\mathcal{P}_4)$ are NP-hard by focusing on the case where: 1) the measurement matrix C is the identity matrix ($C = I$), 2) the measurement noise v_k is Gaussian with zero mean and covariance the identity matrix ($\mathbb{C}(v_k) = I$), and 3) $K = 0$, in which case $z_K = x_{0:K} = x_0$, and thus, $(\mathcal{P}_1)$–$(\mathcal{P}_4)$ are equivalent to the following problem:

$$\begin{aligned} \underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{minimize}} \quad & \log\det(\Sigma(\hat{x}_0)) \\ \text{subject to} \quad & |\mathcal{S}| \leq r. \end{aligned} \tag{13}$$

In more detail, we prove that the problem in (13) is NP-hard by proving that the following problem is equivalent to (13), and that it is NP-hard:

$$\begin{aligned} \underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{minimize}} \quad & \mathbb{H}(x_0|y_0(\mathcal{S})) \\ \text{subject to} \quad & |\mathcal{S}| \leq r \end{aligned} \tag{14}$$

where $\mathbb{H}(x_0|\mathcal{S})$ is the entropy of x_0 given the measurements $y_0(\mathcal{S})$ collected by the selected sensors in $\mathcal{S}$ at time $k = 0$. Specifically, we prove that the problem in (14) is NP-hard by proving it is equivalent to the entropy maximization problem

$$\begin{aligned} \underset{\mathcal{S}\subseteq\{1,2,\ldots,c\}}{\text{maximize}} \quad & \mathbb{H}(y_0(\mathcal{S})) \\ \text{subject to} \quad & |\mathcal{S}| \leq r \end{aligned} \tag{15}$$

which we prove to have an equivalent instance to the NP-hard instance of the entropy maximization problem in [44].

To prove that the problems in (13) and (14) are equivalent, we use [34, Proposition 2], which implies for $K = 0$ that

$$\mathbb{H}(x_0|y_0(\mathcal{S})) = \frac{\log\det(\Sigma(\hat{x}_0))}{2} + \frac{n\log(2\pi e)}{2} \tag{16}$$

where $\hat{x}_0$ is the minimum mean square estimator of x_0 given y_0 [45, Appendix E]. Therefore, (16) implies that minimizing the objective in (13) is equivalent to minimizing the objective in (14).

We next prove that the problems in (14) and (15) are equivalent. To this end, first observe that

$$\mathbb{H}(x_0|y_0(\mathcal{S})) = \mathbb{H}(x_0) + \mathbb{H}(y_0(\mathcal{S})|x_0) - \mathbb{H}(y_0(\mathcal{S})) \tag{17}$$

(we derive (17) using the conditional entropy chain rule [46]), as well as that

1) $\mathbb{H}(x_0)$ is constant with respect to $\mathcal{S}$; and
2) $\mathbb{H}(y_0(\mathcal{S})|x_0)$ is constant for $C = I$, $\mathbb{C}(v_0) = I$, and for fixed $|\mathcal{S}|$, (which is the case throughout this proof, since due to the monotonicity of the $\log\det$ and the entropy, in all problems in (13), (14), and (15), it is $|\mathcal{S}| = r$ for any $\mathcal{S}$ that solves (13), (14), and (15)), since if $y_0^{(i)}(\mathcal{S})$ ($x_0^{(i)}$, respectively) denotes the ith element of $y_0(\mathcal{S})$ (x_0, respectively), then

$$\begin{aligned} &\mathbb{H}(y_0(\mathcal{S})|x_0) \\ &\overset{a}{=} \sum_{i=1}^{|\mathcal{S}|} \mathbb{H}(y_0^{(i)}(\mathcal{S})|x_0, y_0^{(1)}(\mathcal{S}), \ldots, y_0^{(i-1)}(\mathcal{S})) \\ &\overset{b}{=} \sum_{i\in\mathcal{S}} \mathbb{H}(y_0^{(i)}(\mathcal{S})|x_0^{(i)}) \overset{c}{=} \frac{1}{2}\log(2\pi e)|\mathcal{S}|. \end{aligned}$$

In particular, equalities (a), (b), and (c) hold for the following reasons, respectively: (a) holds due to the conditional entropy chain rule [46]; (b) holds since for all $j \neq i$, $y_0^{(i)}$ given $x_0^{(i)}$ is independent of $x_0^{(j)}$ and $y_0^{(j)}$; and (c) holds since $y_0^{(i)}$ given $x_0^{(i)}$ is Gaussian with variance 1, since we consider the case where $\mathbb{C}(v_0) = I$.

Due to (17) and the latter two observations, we conclude that (14) is equivalent to (15).

Given the equivalence of (14) and (15), we conclude the proof by finding an instance for the problem in (13) that is equivalent to an instance for the problem in (15) that is NP-hard. In particular, consider Σ to be any $n \times n$ matrix that makes the entropy maximization problem in [44, Th. 1] NP-hard: Σ is a positive definite symmetric with all the diagonal entries equal to $3n$, and all the off-diagonal entries equal to 0 or 1. The problem (15) is NP-hard if we can find an instance for the problem in (13) where $y_0(\{1, 2, \ldots, c\})$ is a Gaussian random variable with covariance Σ. Indeed, let $\mathbb{C}(x_0)$ be any positive definite symmetric matrix with all diagonal entries equal to $3n - 1$, and all off-diagonal entries equal to 0 or 1 (Σ_0 is positive define by construction, since it is both diagonally dominant, and as a result invertible, and symmetric). For this selection of parameters, y_0 has covariance Σ; the reason is threefold: 1) $y_0 = x_0 + v_0$, 2) x_0 and v_0 are Gaussian with covariances Σ_0 and I, respectively, and 3) x_0 and v_0 are uncorrelated; as a result, y_0 is Gaussian with covariance $\Sigma_0 + I = \Sigma$. ■

Proof of Theorem 2: In the following paragraphs, we only present the proof for $\log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right)$, since the proof for $\log\det\left(\Sigma_{\hat{x}_{0:K}}(\mathcal{S})\right)$ follows similar steps. In particular, we complete the proof for $\log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right)$ in three steps: 1) we prove its closed formula; 2) we show that it is nonincreasing; and 3) we prove that it is supermodular.

Given any finite estimation horizon $[0, K]$, we first prove that

$$\begin{aligned} &\log\det\left(\Sigma_{\hat{z}_K}(\mathcal{S})\right) \\ &\quad = -\log\det\left(\sum_{i=1}^{c} s_i L_{0:K}^\top M^{(i)} L_{0:K} + \mathbb{C}(z_K)^{-1}\right). \end{aligned} \tag{18}$$

In particular, as in [33, Proof of Lemma 1], we let $S_{0:K}$ denote the block diagonal matrix with diagonal elements $K + 1$ copies of the sensor placement matrix S in (8). Then, using the Woodbury matrix identity [47, Corollary 2.8.8] at (3), we obtain $\Sigma_{\hat{z}_K}(\mathcal{S}) = (\mathcal{O}_K^\top \Xi \mathcal{O}_K + \mathbb{C}(z_K)^{-1})^{-1}$, where $\Xi \equiv (S_{0:K}\mathbb{C}(v_{0:K})S_{0:K}^\top)^{-1}$, and due to the definitions of $S_{0:K}$, $C_{0:K}$

and $\mathcal{O}_K = S_{0:K} C_{0:K} L_{0:K}$. Moreover, due to the definition of $S_{0:K}$, i.e., it contains a block of matrices that are only zero or identity matrices, and because $\mathbb{C}(v_{0:K})$ is a block diagonal, it follows that $\Xi = S_{0:K}\mathbb{C}(v_{0:K})^{-1}S_{0:K}^\top$, which can be verified by direct calculation. Overall, we obtain

$$\Sigma_{\hat{z}_K}(\mathcal{S}) = (L_{0:K}^\top C_{0:K}^\top \Lambda \mathbb{C}(v_{0:K})^{-1}\Lambda C_{0:K} L_{0:K} + \mathbb{C}(z_K)^{-1})^{-1}$$

where $\Lambda \equiv S_{0:K}^\top S_{0:K}$. Now, since Λ and $\mathbb{C}(v_{0:K})^{-1}$ are block diagonal, and the blocks of Λ are either identity or zero matrices, $\mathbb{C}(v_{0:K})^{-1}\Lambda = \Lambda\mathbb{C}(v_{0:K})^{-1}$. Furthermore, the definition of $S_{0:K}$ implies that $\Lambda^2 = \Lambda$. Thus, it follows that:

$$\Sigma_{\hat{z}_K}(\mathcal{S}) = (L_{0:K}^\top C_{0:K}^\top \Lambda \mathbb{C}(v_{0:K})^{-1} C_{0:K} L_{0:K} + \mathbb{C}(z_K)^{-1})^{-1}. \tag{19}$$

For the last step, observe first that $\Lambda = \sum_{i=1}^c s_i I^{(i)}$, so

$$\begin{aligned} &L_{0:K}^\top C_{0:K}^\top \Lambda \mathbb{C}(v_{0:K})^{-1} C_{0:K} L_{0:K} \\ &\quad= \sum_{i=1}^c s_i L_{0:K}^\top C_{0:K}^\top I^{(i)} \mathbb{C}(v_{0:K})^{-1} C_{0:K} L_{0:K} \qquad (20) \\ &\quad= \sum_{i=1}^c s_i L_{0:K}^\top M^{(i)} L_{0:K} \qquad (21) \end{aligned}$$

where we derive (21) from (20) by using for $I^{(i)}$ the reverse steps to the ones we used for Λ to derive (19).

Next, to prove that $\log\det(\Sigma_{\hat{z}_K}(\mathcal{S}))$ is a nonincreasing set function in the choice of the sensors $\mathcal{S}$, we follow similar steps to those in [33, Th. 2]. Specifically, consider $\mathcal{S} \subseteq \mathcal{S}'$, and observe that (18) and from [47, Th. 8.4.9], $\Sigma_{\hat{z}_K}(\mathcal{S}') \preceq \Sigma_{\hat{z}_K}(\mathcal{S})$, since $L_{0:K}^\top M^{(i)} L_{0:K} \succeq 0$ and $\mathbb{C}(z_K) \succ 0$. As a result, $\log\det(\Sigma_{\hat{z}_K}(\mathcal{S}')) \leq \log\det(\Sigma_{\hat{z}_K}(\mathcal{S}))$.

Finally, to prove that $\log\det(\Sigma_{\hat{z}_K}(\mathcal{S}))$ is a supermodular set function, we prove that $-\log\det(\Sigma_{\hat{z}_K}(\mathcal{S}))$ is submodular. In particular, recall that a function $h : 2^{[c]} \mapsto \mathbb{R}$ is submodular if and only if, for any $a \in [c]$, the function $h_a : 2^{[c]\setminus\{a\}} \mapsto \mathbb{R}$, where $h_a(\mathcal{S}) \equiv h(\mathcal{S}\cup\{a\}) - h(\mathcal{S})$, is a nonincreasing set function. Therefore, to prove that $h(\mathcal{S}) = -\log\det(\Sigma_{\hat{z}_K}(\mathcal{S}))$ is submodular, we may prove that the $h_a(\mathcal{S})$ is a nonincreasing set function. To this end, we denote $\sum_{i=1}^c s_i L_{0:K}^\top M^{(i)} L_{0:K}$ in (18) by $M(\mathcal{S})$, and follow similar steps to those in [25, Proof of Theorem 6]. Specifically, we note that

$$\begin{aligned} h_a(\mathcal{S}) &= \log\det(M(\mathcal{S}\cup\{a\}) + \mathbb{C}(z_K)^{-1}) \\ &\quad - \log\det(M(\mathcal{S}) + \mathbb{C}(z_K)^{-1}) \\ &= \log\det(M(\mathcal{S}) + M(\{a\}) + \mathbb{C}(z_K)^{-1}) \\ &\quad - \log\det(M(\mathcal{S}) + \mathbb{C}(z_K)^{-1}). \end{aligned}$$

For $\mathcal{S} \subseteq \mathcal{S}'$ and $t \in [0,1]$, define $\Phi(t) \equiv \mathbb{C}(z_K)^{-1} + M(\mathcal{S}) + t(M(\mathcal{S}') - M(\mathcal{S}))$ and

$$g(t) \equiv \log\det(\Phi(t) + M(\{a\})) - \log\det(\Phi(t)).$$

Then, $g(0) = h_a(\mathcal{S})$ and $g(1) = h_a(\mathcal{S}')$. Moreover, since

$$\frac{d\log\det(\Phi(t))}{dt} = \mathrm{tr}\left(\Phi(t)^{-1}\frac{d\Phi(t)}{dt}\right)$$

(as in [48, eq. (43)]), it follows that:

$$\dot{g}(t) = \mathrm{tr}\left[((\Phi(t) + M(\{a\}))^{-1} - \Phi(t)^{-1})F\right]$$

where $F \equiv M(\mathcal{S}') - M(\mathcal{S})$. Furthermore, from [47, Proposition 8.5.5], we have that

$$(\Phi(t) + M(\{a\}))^{-1} - \Phi(t)^{-1} \preceq 0$$

where $\Phi(t)$ is invertible since $\mathbb{C}(z_K)^{-1} \succ 0$, $M(\mathcal{S}) \succeq 0$, and $M(\mathcal{S}') \succeq M(\mathcal{S})$). Since also $F \succeq 0$, from [47, Corollary 8.3.6], it readily follows that:

$$\lambda_{\max}[((\Phi(t) + M(\{a\}))^{-1} - \Phi(t)^{-1})F] \leq 0.$$

Thus, $\dot{g}(t) \leq 0$, and $h_a(\mathcal{S}') = g(1) = g(0) + \int_0^1 \dot{g}(t)dt \leq g(0) = h_a(\mathcal{S})$, i.e., h_a is nonincreasing. ■

Proof of Theorem 3: The proof of the theorem in attained in three main steps: 1) we prove (9); 2) we prove (10); and 3) we prove the computational complexity of Algorithm 1.

To prove (9), let $\mathcal{S}_0, \mathcal{S}_1, \ldots$, be the sequence of sets selected by Algorithm 1 and l the smallest index such that $\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}_l) \leq R$. Therefore, $\mathcal{S}_l$ is the set that Algorithm 1 returns. To prove (9), we first observe that Theorem 2 implies that $\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}))$ is a supermodular and nonincreasing. Then, from [49], we have that

$$\frac{l}{|\mathcal{S}^\star|} \leq 1 + \log\frac{\log\det(\Sigma_{\hat{z}_K}, \emptyset) - \log\det(\Sigma_{\hat{z}_K}, [c])}{\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}_{l-1}) - \log\det(\Sigma_{\hat{z}_K}, [c])}.$$

Now, l is the first time that $\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}_l) \leq R$, and as a result $\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}_{l-1}) > R$, and, as a consequence, we have that (10) holds. Furthermore, $\log\det(\Sigma_{\hat{z}_K}, \emptyset) = \log\det(\mathbb{C}(z_K))$, and from the geometric-arithmetic mean inequality, we obtain that

$$\begin{aligned} \log\det(\mathbb{C}(z_K)) &\leq n(K+1)\log\frac{\mathrm{tr}(\mathbb{C}(z_K))}{n(K+1)} \\ &\leq n(K+1)\log\frac{n(K+1)\max(\sigma_0^2, \sigma_w^2)}{n(K+1)} \\ &= n(K+1)\log\max(\sigma_0^2, \sigma_w^2). \end{aligned}$$

Finally, to prove the computational complexity of Algorithm 1, note that the `while` loop is repeated for at most c times. Moreover, the complexity to compute the determinant of an $n(K+1)\times n(K+1)$ matrix, using the Coppersmith–Winograd algorithm [50], is $O((nK)^{2.4})$, which is also the complexity incurred by the multiplication between such two matrices. Additionally, the determinant of at most $c+1$ matrices must be computed so that the

$$\arg\max_{a\in[c]\setminus\mathcal{S}} (\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}) - \log\det(\Sigma_{\hat{z}_K}, \mathcal{S}\cup\{a\}))$$

can be computed. Also, $O(c)$ time is required to find a maximum element between c available. Therefore, the overall computational complexity of Algorithm 1 is dominated by $O(c^2(nK)^{2.4})$. ■

Proof of Theorem 4: In the following paragraphs, we complete the proof of the theorem in three steps: 1) we prove (11); 2) we prove that the approximation factor $1 - 1/e$ in (11) is the best one can achieve in polynomial time for $(\mathcal{P}_2)/(\mathcal{P}_4)$; and 3) we discuss the computational complexity of the algorithm.

To prove (11), we first observe that Theorem 2 implies that $\log\det(\Sigma_{\hat{z}_K}, \mathcal{S}) - \log\det(\Sigma_{\hat{z}_K}, \emptyset)$ is a supermodular,

nonincreasing, and nonpositive set function. Consequently, the results from [49] can be invoked to obtain (11).

To prove that the approximation factor $1-1/e$ in (11) is the best one can achieve in polynomial time for $(\mathcal{P}_2)/(\mathcal{P}_4)$, we recall that in the worst case, $(\mathcal{P}_2)/(\mathcal{P}_4)$ are equivalent to the minimal observability problem (see proof of Theorem 1). Then, the result follows by noticing that the minimal observability problem has the same computational complexity as the set cover problem [22], which cannot be approximated in polynomial time with a factor better than $1-1/e$ [40].

Finally, the computational complexity of Algorithm 2 can be derived by following the same steps and reasoning as the one proposed in the proof of Theorem 3 to show the time complexity of Algorithm 1. ■

Proof of Theorem 5: Since the arithmetic mean of a finite set of positive numbers is at least as large as their harmonic mean, the following inequality holds:

$$\text{tr}(\Sigma_{\hat{z}_K}) \geq \frac{(n(K+1))^2}{\text{tr}\left(\sum_{i=1}^{c} s_i L_{0:K}^{\top} M^{(i)} L_{0:K} + \mathbb{C}(z_K)^{-1}\right)} \tag{22}$$

where we used the closed form for $\Sigma_{\hat{z}_K}$ proved in Theorem 2.

Furthermore, in the denominator of (22), for the first term it is $\text{tr}(\sum_{i=1}^{c} s_i L_{0:K}^{\top} M^{(i)} L_{0:K}) = \sum_{i=1}^{c} s_i \text{tr}(M^{(i)} L_{0:K} L_{0:K}^{\top})$, where $\text{tr}(L_{0:K}^{\top} M^{(i)} L_{0:K}) \leq n(K+1)\|\bar{C}\|_2^2 \|\mathbb{C}(v_{1:K})^{-1}\|_2 \|L_{0:K} L_{0:K}^{\top}\|_2$, since $\|I^{(i)}\|_2 = 1$, and for the second term it is $\text{tr}(\mathbb{C}(z_K)^{-1}) \leq n(K+1)\max\{\sigma_0^{(-1)}, \sigma_w^{(-1)}\}$. Therefore,

$$\text{tr}\left(\sum_{i=1}^{c} s_i L_{0:K}^{\top} M^{(i)} L_{0:K} + \mathbb{C}(z_K)^{-1}\right) \leq$$

$$ern(K+1)\sigma_v^{(-1)}\|\bar{C}\|_2^2\|\bar{L}\|_2 + n(K+1)\max\{\sigma_0^{(-1)}, \sigma_w^{(-1)}\}.$$

Hence, $\text{tr}(\Sigma_{\hat{z}_K}) \geq n(K+1)/(r\sigma_v^{(-1)}\|\bar{C}\|_2^2\|\bar{L}\|_2 + \max\{\sigma_0^{(-1)}, \sigma_w^{(-1)}\})$. ■

REFERENCES

[1] T. McMillen, T. Williams, and P. Holmes, "Nonlinear muscles, passive viscoelasticity and body taper conspire to create neuromechanical phase lags in anguilliform swimmers," *PLoS Comput. Biol.*, vol. 4, no. 8, 2008, Art. no. e1000157.

[2] Y. Kobayashi, H. Watanabe, T. Hoshi, K. Kawamura, and M. G. Fujie, "Viscoelastic and nonlinear liver modeling for needle insertion simulation," in *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*. New York, NY, USA: Springer, 2012, pp. 41–67.

[3] C. Wex, M. Fröhlich, K. Brandstädter, C. Bruns, and A. Stoll, "Experimental analysis of the mechanical behavior of the viscoelastic porcine pancreas and preliminary case study on the human pancreas," *J. Mech. Behav. Biomed. Mater.*, vol. 41, pp. 199–207, 2015.

[4] K. Wang, R. McCarter, J. Wright, J. Beverly, and R. Ramirez-Mitchell, "Viscoelasticity of the sarcomere matrix of skeletal muscles. the titin-myosin composite filament is a dual-stage molecular spring," *Biophys. J.*, vol. 64, no. 4, pp. 1161–1177, 1993.

[5] T. M. Best, J. McElhaney, W. E. Garrett, and B. S. Myers, "Characterization of the passive responses of live skeletal muscle using the quasi-linear theory of viscoelasticity," *J. Biomech.*, vol. 27, no. 4, pp. 413–419, 1994.

[6] T. C. Doehring, A. D. Freed, E. O. Carew, and I. Vesely, "Fractional order viscoelasticity of the aortic valve cusp: An alternative to quasilinear viscoelasticity," *J. Biomech. Eng.*, vol. 127, no. 4, pp. 700–708, 2005.

[7] E. Macé, I. Cohen, G. Montaldo, R. Miles, M. Fink, and M. Tanter, "In vivo mapping of brain elasticity in small animals using shear wave imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 550–558, Mar. 2011.

[8] S. Nicolle, L. Noguer, and J.-F. Palierne, "Shear mechanical properties of the spleen: Experiment and analytical modelling," *J. Mech. Behav. Biomed. Mater.*, vol. 9, pp. 130–136, 2012.

[9] N. Grahovac and M. Žigić, "Modelling of the hamstring muscle group by use of fractional derivatives," *Comput. Math. Appl.*, vol. 59, no. 5, pp. 1695–1700, 2010.

[10] V. E. Tarasov, *Fractional Dynamics: Applications of Fractional Calculus to Dynamics of Particles, Fields and Media*. Berlin, Germany: Springer Science & Business Media, 2011.

[11] P. Bogdan, B. M. Deasy, B. Gharaibeh, T. Roehrs, and R. Marculescu, "Heterogeneous structure of stem cells dynamics: Statistical models and quantitative predictions," *Sci. Rep.*, vol. 4, 2014, Art. no. 4826.

[12] M. Ghorbani and P. Bogdan, "A cyber-physical system approach to artificial pancreas design," in *Proc. 9th IEEE/ACM/IFIP Int. Conf. Hardware/Softw. Codesign Syst. Synthesis*, 2013, p. 17.

[13] Y. Xue, S. Rodriguez, and P. Bogdan, "A spatio-temporal fractal model for a CPS approach to brain-machine-body interfaces," in *Proc. IEEE Des. Autom. Test Eur. Conf. Exhib.*, 2016, pp. 642–647.

[14] Y. Xue, S. Pequito, J. M. Coelho, P. Bogdan, and G. J. Pappas, "Minimum number of sensors to ensure observability of physiological systems: A case study," in *Proc. 54th Annu. Allerton Conf. Commun. Control Comput.*, 2016, pp. 1181–1188.

[15] Y. Xue and P. Bogdan, "Constructing compact causal mathematical models for complex dynamics," in *Proc. 8th Int. Conf. Cyber-Phys. Syst.*, 2017, pp. 97–107.

[16] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, pp. 392–395, 2005.

[17] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, "Skeleton and fractal scaling in complex networks," *Phys. Rev. Lett.*, vol. 96, no. 1, 2006, Art. no. 018701.

[18] F. Klimm, D. S. Bassett, J. M. Carlson, and P. J. Mucha, "Resolving structural variability in network models and the brain," *PLoS Comput. Biol.*, vol. 10, no. 3, 2014, Art. no. e1003491.

[19] Y. Xue and P. Bogdan, "Reliable multi-fractal characterization of weighted complex networks: Algorithms and implications," *Sci. Rep.*, vol. 7, 2017, Art. no. 7487.

[20] J. Klafter, S. Lim, and R. Metzler, *Fractional Dynamics: Recent Advances*. Singapore: World Scientific, 2012.

[21] S. Guermah, S. Djennoune, and M. Bettayeb, "Controllability and observability of linear discrete-time fractional-order systems," *Appl. Math. Comput. Sci.*, vol. 18, no. 2, pp. 213–222, 2008.

[22] A. Olshevsky, "Minimal controllability problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 249–258, Sep. 2014.

[23] S. Pequito, G. Ramos, S. Kar, A. P. Aguiar, and J. Ramos, "The robust minimal controllability problem," *Automatica*, vol. 82, pp. 261–268, 2017.

[24] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, Mar. 2014.

[25] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 91–101, Mar. 2016.

[26] V. Tzoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie, "Minimal actuator placement with bounds on control effort," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 67–78, Mar. 2016.

[27] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, "Spectrum of controlling and observing complex networks," *Nature Phys.*, no. 11, pp. 779–786, 2015.

[28] Y. Zhao and J. Cortes, "Gramian-based reachability metrics for bilinear networks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 620–631, Sep. 2017.

[29] S. Pequito, P. Bogdan, and G. J. Pappas, "Minimum number of probes for brain dynamics observability," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 306–311.

[30] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for optimal filtering of linear dynamical systems: Complexity and approximation," in *Proc. 54th IEEE Conf. Decision Control*, 2015, pp. 5002–5007.

[31] M.-A. Belabbas, "Geometric methods for optimal sensor design," *Proc. Math. Phys. Eng. Sci.*, vol. 472, no. 2185, 2016, Art. no. 20150312.

[32] S. Guermah, S. Djennoune, and M. Bettayeb, "Controllability and observability of linear discrete-time fractional-order systems," *Int. J. Appl. Math. Comput. Sci.*, vol. 18, no. 2, pp. 213–222, Jun. 2008.

[33] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Near-optimal sensor scheduling for batch state estimation: Complexity, algorithms, and limits," in *Proc. 55th IEEE Conf. Decision Control*, 2016, pp. 2695–2702.

[34] V. Tzoumas, N. A. Atanasov, A. Jadbabaie, and G. J. Pappas, "Scheduling nonlinear sensors for stochastic process estimation," in *Proc. Amer. Control Conf.*, 2017, pp. 580–585.

[35] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
[36] S. Venkatesh, *The Theory of Probability: Explorations and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[37] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, vol. 1, 2000.
[38] A. Krause and C. Guestrin, "Beyond convexity: Submodularity in machine learning," *ICML Tutorials*, 2008.
[39] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Oper. Res.*, vol. 3, no. 3, pp. 177–188, 1978.
[40] U. Feige, "A threshold of ln(n) for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, 1998.
[41] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[42] "EEG motor movement/imagery dataset," *PhysioNet*. [Online]. Available: http://www.physionet.org/pn4/eegmmidb/ and http://circ.ahajournals.org/cgi/content/full/101/23/e215]
[43] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, and T. Dutoit, "Performance of the Emotiv EPOC headset for P300-based applications," *Biomed. Eng. OnLine*, vol. 12, no. 1, p. 56, Jun. 2013, doi: 10.1186/1475-925X-12-56.
[44] C.-W. Ko, J. Lee, and M. Queyranne, "An exact algorithm for maximum entropy sampling," *Oper. Res.*, vol. 43, no. 4, pp. 684–691, 1995.
[45] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Vol. I, 3rd ed. Belmont, MA, USA: Athena Scientific, 2005.
[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[47] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[48] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Kgs. Lyngby, Denmark: Technical Univ. Denmark, 2012.
[49] L. A. Wolsey, "An analysis of the greedy algorithm for the submodular set covering problem," *Combinatorica*, vol. 2, no. 4, pp. 385–393, 1982.
[50] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," in *Proc. 19th Annu. ACM Symp. Theory Comput.*, 1987, pp. 1–6.

Vasileios Tzoumas (S'12) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2012, the Master of Science degree in electrical engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2016, and the Master of Arts degree in statistics from the Wharton School of Business University of Pennsylvania, in 2016. He is currently working toward the Ph.D. degree at the Department of Electrical and Systems Engineering, University of Pennsylvania, and a visiting Ph.D. student at the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA.

His current research interests include the resource-constrained and resilient design of large-scale systems, with applications in estimation and control; to this end, he uses and develops tools in the theory of submodular function maximization.

Dr. Vasileios was a Best Student Paper Award finalist in the 56th IEEE Conference in Decision and Control in 2017.

Yuankun Xue (S'12) received the B.Sc. and M.Sc. degrees from Fudan University, Shanghai, China, in 2007 and 2011, respectively. He is currently working toward the Ph.D. degree in electrical engineering under the supervision of Professor Paul Bogdan with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA.

His research interests include thoeretical and statistical machine learning, analysis, and control of cyber-physical systems, large-scale dynamic networked systems modeling, optimization and control, and design methodologies for high-performance manycore platforms.

Sérgio Pequito (S'09–M'14) received the B.Sc. and M.Sc. degrees in applied mathematics from the Instituto Superior Tcnico, Lisbon, Portugal, in 2007 and 2009, respectively. He received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, and Instituto Superior Técnico, through the CMU-Portugal Program, in 2014.

He is an Assistant Professor with the Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute. From 2014 to 2017, he was a Postdoctoral Researcher in General Robotics, Automation, Sensing and Perception Laboratory (GRASP Lab), University of Pennsylvania. His research consists of understanding the global qualitative behavior of large-scale systems from their structural or parametric descriptions and provide a rigorous framework for the design, analysis, optimization and control of large scale (real-world) systems. Currently, his interests span to neuroscience and biomedicine, where dynamical systems and control theoretic tools can be leveraged to develop new analysis tools for brain dynamics that, ultimately, will lead to new diagnostics and treatments of neural disorders. In addition, these tools can be used towards effective personalized medicine and improve brain-computer and brain-machine-brain interfaces that will improve people's life quality.

Dr. Pequito was the best student paper finalist in the 48th IEEE Conference on Decision and Control in 2009. He received the ECE Outstanding Teaching Assistant Award from the Electrical and Computer Engineering Department, Carnegie Mellon University, and the Carnegie Mellon Graduate Teaching Award (University-wide) honorable mention, both in 2012. Also, he was a 2016 EECI European Ph.D. Award on Control for Complex and Heterogeneous Systems finalist and received the 2016 O. Hugo Schuck Award in the Theory Category.

Paul Bogdan (M'11) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2012.

He is currently an Assistant Professor with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA. His current research interests include theoretical foundations of cyber-physical systems, control of complex time-varying interdependent networks, modeling and analysis of biological systems and swarms, new control algorithms for dynamical systems exhibiting multi-fractal characteristics, modeling biological /molecular communication, fractal mean field games, performance analysis and design methodologies for many core systems.

George J. Pappas (F'09) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 1998.

He is currently the Joseph Moore Professor and the Chair with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. He also holds a secondary appointment with the Department of Computer and Information Sciences and the Department of Mechanical Engineering and Applied Mechanics. He is a member of the GRASP Lab and the PRECISE Center. He had previously served as the Deputy Dean for Research with the School of Engineering and Applied Science. His research interests include control theory and, in particular, hybrid systems, embedded systems, cyberphysical systems, and hierarchical and distributed control systems, with applications to unmanned aerial vehicles, distributed robotics, green buildings, and biomolecular networks.

Dr. Pappas has received various awards, such as the Antonio Ruberti Young Researcher Prize, the George S. Axelby Award, the Hugo Schuck Best Paper Award, the George H. Heilmeier Award, the National Science Foundation PECASE Award, and numerous best student papers awards.