REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
18-08-2022	Final Report	23-Jan-2018 - 22-Feb-2022

4. TITLE AND SUBTITLE
Final Report: Effective Control of Leader-Follower Networks

5a. CONTRACT NUMBER
W911NF-18-1-0072

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Boston University
Office of Sponsored Program
881 Commonwealth Avenue
Boston, MA 02215 -1300

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
72590-NS.6

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Alexander Olshevsky
UU	UU	UU	UU		19b. TELEPHONE NUMBER
					617-353-0173

RPPR Final Report
as of 23-Aug-2022

Agency Code: 21XD

Proposal Number: 72590NS **Agreement Number: W911NF-18-1-0072**
INVESTIGATOR(S):

Name: Alexander Olshevsky
Email: alexols@bu.edu
Phone Number: 6173530173
Principal: Y

Organization: **Boston University**
Address: Office of Sponsored Program, Boston, MA 022151300
Country: USA
DUNS Number: 049435266 EIN: 042103547
Report Date: 22-May-2022 Date Received: 18-Aug-2022
Final Report for Period Beginning 23-Jan-2018 and Ending 22-Feb-2022
Title: Effective Control of Leader-Follower Networks
Begin Performance Period: 23-Jan-2018 **End Performance Period:** 22-Feb-2022
Report Term: 0-Other
Submitted By: Alexander Olshevsky Email: alexols@bu.edu
Phone: (617) 353-0173

STEM Degrees: **STEM Participants:**

Major Goals: The main goal of this project is to develop a theory of input selection for effective control of linear and some nonlinear systems. In contrast to traditional approaches in control theory, which often take both the system and the set of actuated variables as fixed, our starting point will be only the system itself. We will then develop algorithms to decide which variables of the system should be affected with an input to optimize a variety of control objectives.

Actuator selection problems appear throughout control and engineering, but the primary motivating application for this work is multi-agent control. It is expected that future military missions will be performed in part by autonomous or semi-autonomous robotic platforms. It is often undesirable to give each drone in a swarm its own commands and an alternative approach is to have a single node be controlled with the remaining nodes using a leader-following method. Algorithm for deciding which variables to actuate can be used to decide which nodes of the swarm should be controlled.

This overarching goal will be broken up into a number of thrusts as follows.

1. Develop efficient methods for quantifying the number of controlled ``leader'' nodes needed to achieve controllability of a multi-agent network. In particular, in multi-agent networks our goal is to design algorithms for leader selection, where leaders will be controlled by a specially designated input signal, so that the entire graph topology (consisting of all the agents) can be efficiently controlled through a process where all the non-leader nodes follow the strategies of their neighbors and of the leaders.

2. Develop measures of "control importance" quantifying the importance of each node for controllability objectives. These measures will act as heuristics for leader placement, allowing us to choose specific nodes in the network which should be chosen as leaders for effective control of multi-agent flocks to take place.

3. Derive impossibility results for effective control which will act as design guidelines. Specifically, networks which cannot be controlled efficiently, either due to computational infeasibility or the unavoidable necessity of large inputs for effective control, should be classified.

4. Analyze the energy expenditure metric in optimal control in the context of actuator selection. It is known that optimizing controllability often results in practically unrealistic
designs which use extremely large inputs in optimal control. Thus bounds on energy expenditure need to be

considered explicitly in the problem formulation in order to produce realistic input selection. Unfortunately, this leads to a variety of non-convex problems, and understanding both how tractable these problems are and how to relax them will be a goal of this project.

5. The duality of controllability and observability means that the advances above are closely related to problems in network observability and estimation. Thus this project will also consider "leader selection" problems dealing with understanding which nodes of a network should make observations, and which should act passively in merely relaying messages, provided that the number of sensing nodes which can make observation is bounded, or provided that observations are costly.

Accomplishments: (1) We provided an efficient computational method to solve an integer relaxation of the leader selection problem. Specifically, the goal was to control a linear system x'=Ax by choosing to actuate a small set of variables so that the resulting system with input x'=Ax+Bu is not only controllable, but one can move it using optimal control without spending too much energy. Unfortunately, this problem formulation leads to the optimization of the inverse of the smallest eigenvalue of the inverse of the controllability Grammian, which is notoriously non-convex, and for which a number of intractability results exist.

A natural idea is to maximize the trace of the controllability Grammian, rather than minimizing its inverse. The solution to this relaxed problem provides approximation guarantees for the original problem and the objective of the relaxed problem is convex. A difficulty is that, even with a convex objective, it is not clear that the problem can be solved efficiently, because the input selection problem is inherently an integer (or, rather, binary) programming problem: for each possible input location, we must select a binary variable {0,1} capturing whether we intend to put an actuator in that location.

The starting point of our work is a result produced by Ikeda and Kashima which won the inaugural award for being the best paper published at the IEEE Conference on Decision and Control in 2018 & 2019. Ikeda & Kashima gave a condition for when the problem is equivalent to a relaxed version with continuous variables in [0,1]. The condition needed by Ikeda & Kashima for this involved the non-constancy of certain functions coming from the controllability Grammian.

We proved that the result of Ikeda and Kashima -- that one can solve the [0,1]-relaxed problem and obtain a solution of the binary {0,1} problem -- holds with no conditions whatsoever, as long as the matrices are nonzero. In particular, the relaxed problem of maximizing the trace of the controllability Grammian can always be solved.

(2) We obtained an intractability result dealing with the (in)approximability of the Witsenhausen problem, a classic benchmark in decentralized control. In the Witsenhausen problem, two agents act sequentially to control a linear system, with no information flowing from the first agent to the second agent. The notoriety of the Witsenhausen problem comes from the many attempts to solve it since its introduction in the 1970s, all without success. We showed that the Witsenhausen problem is intractable to approximately to a multiplicative factor that grows quadratically in the size of the input distribution. One might argue this result explains why so little progress on the Witsenhausen problem has been made: previous generations of researchers in the control community were attempting to tackle a problem which is, in general, intractable. This result has implications for sequential control of leader-follower networks. In contrast to the situation where all the leaders and followers act in tandem, in the case where the leaders act first, optimal strategies may be impossible to compute. This result can thus be used as an implicit justification of using non-sequential models in control of networks.

(3) An impossibility result was proven for the problem of selecting actuators/leaders to make a linear system controllable over a given subspace. We have shown that there is no better algorithm than simply actuating all the variables (i.e., every node has to be a leader). This is important for two reasons. It is counterintuitive because it is very different from the problem of actuator selection to make the system controllable (for which good algorithms exists that do *not* actuate all the variables). More importantly, this finding contradicts several papers in the literature that claimed otherwise.

(4) A new approach to actuator selection using randomized sampling was developed, with near optimal results for the problem of time-varying actuator selection for minimizing energy expenditure. The core of the approach is an analogy between graph sparsification and actuator selection.

The main contribution of our work was to draw out the analogy between these two problems and find a control-theoretic analogue of the graph resistance in terms of the controllability Grammian. We develop this analogue and

propose to use it as a guide to sampling rows and columns of the Grammian, i.e., sampling actuators. The punchline is that we can solve the problem of time-varying actuator selection to minimize control energy over a time horizon T to any epsilon accuracy in polynomial time -- provided the time horizon T scales at least as (n log n) /epsilon^2, where n is the underlying dimension of the system. This essentially solves a problem posed by Michael Athans in 1972, testifying to the long history of this problem within the control field.

(5) Progress was made on the problem of cooperative learning in a network: a collection of network nodes are collecting measurements about an unknown state of the world, with the goal being to correctly identify such an unknown state. The focus is usually on the situation when no node alone can accurately identify the state of the world but this is possible if nodes cooperate.

As a simple example, consider the problem of localization: a target is located at an unknown location, and sensors can measure the distance to that location. This is a plausible model when, for example, the target is being tracked by measuring the strength of a wireless signal it is emitting; the strength of the signal only allows you to estimate distance.

We have obtained a formula for the asymptotic tracking error which results in measurements by individual nodes and fusion through a gossiping strategy. A central question here is leader selection: how does the tracking or estimation error depend on which subset of the nodes is doing the measurement? How does it depend on the position of the nodes performing the measurement in the graph topology?

One of the major innovations of this work related to previous work by the PI and others is that we are able to handle when the number of hypotheses is infinite. The reason this is important is that the number of hypotheses typically is infinite -- for example, in the localization problem one can think of each location of the target as a hypothesis. Furthermore, previously proposed algorithms proceeded by communicating vectors between nodes that have the same dimensionality as the total number of hypotheses. If a problem with an infinite number of hypotheses is discretized to a finite number, the number of communications among agents will need to go to infinity to achieve small error in the discretization.

In our work, communication among agents for target tracking scales with the intrinsic dimensionality of the problem instead of the total number of hypotheses. We show that the number of inter-agent communications per step when the distributions come from an exponential family is proportional only to the dimensionality of this exponential family. Further, we give a formula for the convergence rate to the optimal hypothesis. This is done in terms of the "densities" of competing hypotheses as they approach the optimal hypotheses. Surprisingly, we find that the asymptotic decay of the tracking error depends only on the errors of the individual nodes who are doing the measuring, and is not dependent on their positions in the graph topology.

Training Opportunities: Several students were funded by this award and the PI spent time training them to present these results to scientific audiences.

RPPR Final Report
as of 23-Aug-2022

Results Dissemination: All of the papers listed below were put on the arxiv before submission, so that public versions of these results are accessible.

1. Non-asymptotic Concentration Rates in Cooperative Learning Part I: Variational Non-Bayesian Social Learning,
C. Uribe, A. Olshevsky, A. Nedic
IEEE Transactions on Control of Network Systems, 2022

2. Non-asymptotic Concentration Rates in Cooperative Learning Part II: Inference on Compact Hypothesis Sets,
C. Uribe, A. Olshevsky, A. Nedic
IEEE Transactions on Control of Network Systems, 2022

3. Deterministic and Randomized Actuator Scheduling With Guaranteed Performance Bounds,
M. Siami, A. Olshevsky, A. Jadbabaie,
IEEE Transactions on Automatic Control, 2021.

4. On A Relaxation of Time-Varying Actuator Placement,
A. Olshevsky,
IEEE Control System Letters, 2020

5. On The Inapproximability of the Discrete Witsenhausen Problem,
A. Olshevsky,
IEEE Control System Letters, 2019

6. Minimal Reachability is Hard To Approximate
A. Jadbabaie, A. Olshevsky, G. J. Pappas, V. Tzoumas
IEEE Transactions on Automatic Control, vol. 64, no. 2, 2019

7. On (non) Supermodularity of Control Energy
A. Olshevsky,
IEEE Transactions on Control of Network Systems, 2018.

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI
Participant: Alexander Olshevsky
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Rui Liu
Person Months Worked: 6.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Artin Spiridonoff

RPPR Final Report
as of 23-Aug-2022

Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Qianqian Ma
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Control of Network Systems
Publication Identifier Type: DOI Publication Identifier: 10.1109/TCNS.2017.2691463
Volume: 5 Issue: 3 First Page #: 1177
Date Submitted: 9/24/18 12:00AM Date Published: 9/1/18 4:00AM
Publication Location:
Article Title: On (Non)Supermodularity of Average Control Energy
Authors: Alex Olshevsky
Keywords: control-energy network-control
Abstract: Given a linear system, we consider the expected energy to move from the origin to a uniformly random point on the unit sphere as a function of the set of actuated variables. We prove that this function is not necessarily supermodular, correcting some claims in the existing literature.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 2-Awaiting Publicat
Journal: IEEE Transactions on Automatic Control
Publication Identifier Type: DOI Publication Identifier: 10.1109/TAC.2020.3000976
Volume: Issue: First Page #: 1
Date Submitted: 8/10/20 12:00AM Date Published:
Publication Location:
Article Title: Deterministic and Randomized Actuator Scheduling With Guaranteed Performance Bounds
Authors: Milad Siami, Alexander Olshevsky, Ali Jadbabaie
Keywords: control of networks
Abstract: In this paper, we investigate the problem of actuator selection for linear dynamical systems. We develop a framework to design a sparse actuator schedule for a given large-scale linear system with guaranteed performance bounds using deterministic polynomial-time and randomized approximately linear-time algorithms. First, we introduce systemic controllability metrics for linear dynamical systems that are monotone and homogeneous with respect to the controllability Gramian. We show that several popular and widely used optimization criteria in the literature belong to this class of controllability metrics. Our main result is to provide a polynomial-time actuator schedule that on average selects only a constant number of actuators at each time step, independent of the dimension, to furnish a guaranteed approximation of the controllability metrics in comparison to when all actuators are in use. Our results naturally apply to the dual problem of sensor selection, in which we provide a guara
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 23-Aug-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Automatic Control
Publication Identifier Type: DOI Publication Identifier: 10.1109/TAC.2018.2836021
Volume: Issue: First Page #: 1
Date Submitted: 8/10/20 12:00AM Date Published:
Publication Location:
Article Title: Minimal Reachability is Hard To Approximate
Authors: Ali Jadbabaie, Alexander Olshevsky, George J. Pappas, Vasileios Tzoumas
Keywords: Approximation algorithms, computational complexity, controllability, (non-)submodularity, sparse actuator placement
Abstract: In this note, we consider the problem of choosing,which nodes of a linear dynamical system should be actuated so that the state transfer from the system's initial condition to a given final state is possible. Assuming a standard complexity hypothesis,we show that this problem cannot be efficiently solved or approximated in polynomial, or even quasi-polynomial, time
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Control Systems Letters
Publication Identifier Type: DOI Publication Identifier: 10.1109/LCSYS.2019.2911925
Volume: 3 Issue: 3 First Page #: 529
Date Submitted: 8/10/20 12:00AM Date Published: 7/1/19 4:00AM
Publication Location:
Article Title: On the Inapproximability of the Discrete Witsenhausen Problem
Authors: Alex Olshevsky
Keywords: Decentralized control, computational complexity
Abstract: In this note, we consider the problem of choosing,which nodes of a linear dynamical system should be actuated so that the state transfer from the system's initial condition to a given final state is possible. Assuming a standard complexity hypothesis,we show that this problem cannot be efficiently solved or approximated in polynomial, or even quasi-polynomial, time
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Control Systems Letters
Publication Identifier Type: DOI Publication Identifier: 10.1109/LCSYS.2020.2990099
Volume: 4 Issue: 3 First Page #: 656
Date Submitted: 8/10/20 12:00AM Date Published: 7/1/20 4:00AM
Publication Location:
Article Title: On a Relaxation of Time-Varying Actuator Placement
Authors: Alex Olshevsky
Keywords: Control of networks, network analysis andcontrol, optimization
Abstract: We consider the time-varying actuator placement in continuous time, where the goal is to maximize the trace of the controllability Grammian. A natural relaxation of the problem is to allow the binary {0, 1} variable indicating whether an actuator is used at a given time to take on values in the closed interval [0, 1]. We show that all optimal solutions of both the original and the relaxed problems can be given via an explicit formula, and that, as long as the input matrix has no zero columns, the solutions sets of the original and relaxed problem coincide
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 23-Aug-2022

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Alexander Olshevsky
Signature Date: 8/18/22 6:05PM

On (Non)Supermodularity of Average Control Energy

Alex Olshevsky

***Abstract*—Given a linear system, we consider the expected energy to move from the origin to a uniformly random point on the unit sphere as a function of the set of actuated variables. We prove that this function is not necessarily supermodular, correcting some claims in the existing literature.**

***Index Terms*—Linear systems, optimization methods.**

I. Introduction

THIS PAPER is concerned with a property of the actuator selection problem. Given the linear system

$$\dot{x}_i = \sum_{j=1}^{n} a_{ij} x_j,\ i = 1, \ldots, n$$

the simplest actuator selection problem asks for the smallest possible set of variables to affect with an input in order to achieve a prespecified control objective. Typical control objectives include controllability of the resulting system or the ability to steer the system subject to an energy constraint.

Formally, if we choose to affect the set of variables $\{x_i \mid i \in I\}$, then the resulting system-with-input is

$$\begin{aligned} \dot{x}_i &= \sum_{j=1}^{n} a_{ij} x_j + u_i,\ i \in I \\ \dot{x}_i &= \sum_{j=1}^{n} a_{ij} x_j,\ i \notin I \end{aligned} \tag{1}$$

and the goal is to choose the set I as small as possible while still satisfying some control objective. More complex versions of actuator selection problem might not allow one to directly affect each variable; rather, one instead assumes that the system can only be affected in several distinct "sites" and affecting each site affects some subset of the variables all at once.

The actuator selection problem received some attention recently (e.g., [1], [3], and [5]), due to the emergence of recent interest in large-scale systems, for example, in power networks or systems biology. It may be impractical or uneconomical to steer large systems by affecting every, or even most, of the variables, and, consequently, it is natural to ask if the system can be efficiently steered by affecting only very few select variables.

Manuscript received September 29, 2016; revised January 31, 2017; accepted March 19, 2017. Date of publication April 5, 2017; date of current version September 17, 2018. This work was supported by the National Science Foundation under Grant ECCS-1351684 and ARO project 0011128922. Recommended by Associate Editor F. Fagnani. *(Corresponding Author: Alexander Olshevsky.)*

The author is with the Department of Electrical and Computer Engineering and Division of Systems Engineering, Boston University, Boston, MA 02215 USA (e-mail: alexols@bu.edu).

Digital Object Identifier 10.1109/TCNS.2017.2691463

A key property for actual selection problems is supermodularity. A formal definition can be found in the next section, but, roughly speaking, this is the property that affecting variables runs into diminishing returns; that is to say, affecting a certain variable has less impact on the control objective if more variables have already been affected.

Supermodularity is important because it can lead to algorithms with rigorous approximation guarantees. For example, an approximate algorithm for actuator selection to render the system controllable based on supermodularity of the dimension of the controllable subspace was given in [1].[1]

Supermodularity of a number of a control objectives was studied in the recent papers [3] and [5]. Specifically, one of the control objectives studied in [3] was the trace of the inverse of the controllability Gramian, which has the interpretation of being proportional to the expected energy to move from the origin to a random point on the unit sphere (we will refer to this as the *average control energy*). It was claimed in [3] that, for a stable system, average control energy is a supermodular function of the set of affected sites. Using similar arguments, the latter paper [5] claimed that (an arbitrarily small perturbation of) average control energy is a supermodular function of the set of affected variables.

The purpose of this note is to rigorously prove that average control energy is not always supermodular, contrary to what is claimed in [3] and [5]. In other words, we give a proof that there exists a (stable, symmetric) linear system and two sets of variables, $I_1 \subset I_2$ such that average control energy decreases *more* when a certain variable is added to the bigger set of actuated variables I_2, as compared to the scenario when the same variable is added to the smaller set I_1.

The remainder of this paper is organized as follows. In Section II, we give the basic definitions used in the remainder of this paper. The subsequent Section III contains the constructions of linear systems for which average control energy is not supermodula. Finally, Section IV concludes with some brief remarks.

A. Notation

We use the standard notation of letting e_i denote the ith basis vector and I_k to denote the $k \times k$ identity matrix. For a matrix M, we will use M' to denote its transpose. The complement of a set S will be denoted by S^c. The notation 1_k will be used for the column vector of all ones in $\mathbb{R}^k$. Finally, a matrix is called strictly stable if all of its eigenvalues have negative real parts.

[1]Note that although [1] did not use the words "supermodularity" or "submodularity," some of the steps of the proofs were formulations of this property.

II. Basic Definitions

A. Average Control Energy of Linear Systems

Given the linear system

$$\dot{x} = Ax + Bu \tag{2}$$

and an initial state x_0 along with a final state $x_{\rm f}$, we define the control energy $\mathcal{E}(A, B, x_0 \to x_{\rm f}, T)$ to be the minimal energy $\int_0^T ||u(t)||_2^2\, dt$ among all inputs $u : [0, T] \to \mathbb{R}$, which result in $x(T) = x_{\rm f}$ starting from $x(0) = x_0$. If there is no input that results in $x(T) = x_f$ when $x(0) = x_0$, we will adopt the convention that $\mathcal{E}(A, B, x_0 \to x_f)$ is infinite.

The quantity $\mathcal{E}(A, B, x_0 \to x_{\rm f})$ measures the difficulty of steering the system from x_0 to x_f; obviously it will depend on both the starting point x_0 and the final point x_f. One way to obtain a measure of the "difficulty of controllability" of the entire system is to consider the energy involved in moving the system from the origin to a uniformly random point on the unit sphere, namely

$$\mathcal{E}_{\rm ave}(A, B, T) := \int_{||y||_2=1} \mathcal{E}(A, B, 0 \to y, T)\, dy.$$

It is easy to see that this quantity can be written in terms of controllability Gramian. Indeed, first, we define the controllability Gramian $W(T)$ in the usual way as

$$W(A, B, T) := \int_0^T e^{At} BB' e^{A't}\, dt \tag{3}$$

where we will allow T to be equal to $+\infty$ with the proviso that $W(+\infty)$ is welldefined only as long as the matrix A is strictly stable. It is then not difficult to see that

$$\mathcal{E}_{\rm ave}(A, B, T) = \frac{1}{n} {\rm tr}\left[W(A, B, T)^{-1}\right].$$

Moreover, if $W(A, B, T)$ is not invertible, then $\mathcal{E}_{\rm ave}(A, B, T)$ is infinite.

B. Actuator Selection Problem

Before giving a formal statement of the actuator selection problem, let us introduce some notation. First, we will need notation for the dimensions of A and B; specifically, let us suppose $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then, given a set $S \subset \{1, \ldots, m\}$, we denote $B(S)$ to be the matrix in $\mathbb{R}^{n\times|S|}$ composed of the columns of B corresponding to indices in S. For example, if $B = I_3$ (the 3×3 identity matrix), then

$$B(\{1, 2\}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We then define

$$\mathcal{E}_{\rm ave}(A, B, T, S) := \mathcal{E}_{\rm ave}(A, B(S), T).$$

In other words, $\mathcal{E}_{\rm ave}(A, B, T, S)$ is the average control energy at time T when using only the columns of B corresponding to the indices in the set S.

There are many versions of actuator selection problems, but the ones we consider here ask to optimize the function $\mathcal{E}_{\rm ave}(A, B, T, S)$ for fixed A, B, T as a function of the set S. For example, a natural problem is to ask for S of smallest cardinality meeting the energy constraint $\mathcal{E}_{\rm ave} \le c$ for some real number c.

In the simplest and most natural case, B is the $n \times n$ identity matrix; in that case, we may think of choosing S as corresponding to actuating the variables of the differential equation $\dot{x} = Ax$ as in (1). More generally, affecting a system in a given "site" may affect a group of variables all-at-once; this is appropriately modeled by a general matrix B, where the choice of adding an index i to S involves affecting all the variables in the ith column of B.

C. Set Functions

A function $f : 2^{\{1,\ldots,m\}} \to \mathbb{R}$ is called nonincreasing if $S_1 \subset S_2$ implies $f(S_1) \ge f(S_2)$. A set function is called supermodular if $S_1 \subset S_2$ and $a \notin S_2$ implies that

$$f(S_1) - f(S_1 \cup \{a\}) \ge f(S_2) - f(S_2 \cup \{a\}). \tag{4}$$

Intuitively, if the function f is supermodular, then adding element a decreases the function less if it is added to the bigger set S_2 as compared to the smaller set S_1.

A set function is called submodular if its negation is supermodular.

III. Average Control Energy May Not Be Supermodular

Throughout this section, we will investigate the setup where A, B, T are fixed and $\mathcal{E}_{\rm ave}(A, B, T, S)$ is considered as a function only of the set S. It is quite easy to see this function is nonincreasing, i.e., average control energy cannot increase when we actuate more places.

As discussed earlier, one might further guess that $\mathcal{E}_{\rm ave}(A, B, T, S)$ would be a supermodular function of S. Indeed, it seems quite intuitive that the gain from actuating any specific variable runs into diminishing returns as other variables become actuated. Strangely enough, it turns out that this intuition is not correct and we now turn to the main point of this note, which is to construct counterexamples for this intuition.

A. (Non)Supermodularity of Average Control Energy for Strictly Stable Matrices

We begin this section with a rigorous proof showing that average control energy may not be supermodular even if the system is strictly stable.

Theorem 1: There exists a 2×2 matrix A and a 2×5 matrix B such that

1) A is strictly stable.
2) $\mathcal{E}_{\rm ave}(A, B(S), +\infty)$ is finite for all nonempty S.
3) $\mathcal{E}_{\rm ave}(A, B(S), +\infty)$ is not a supermodular function of S.

This theorem contradicts [3, Th. 5], which claims that $-\mathcal{E}_{\rm ave}(A, B(S), +\infty)$ is a submodular function of S under the assumptions that 1) $\mathcal{E}_{\rm ave}(A, B(S), +\infty)$ is finite for all S and

2) A is stable. Later in this paper, we will use Theorem 1 to construct a counterexample where A is 6×6 and B is the 6×6 identity matrix.

The error of the proof in [3] is the implicit use of the implication "$U \preceq V$ implies $U^2 \preceq V^2$ for positive semidefinite U, V," which does not hold. The proof of a related assertion in [5] suffers from the same problem.

The proof of Theorem 1, given next, relies primarily on calculation; since the controllability Gramians involved are 2×2, this can be done explicitly without reliance on computer-assisted computations. To find the motivation for the choices made within the course of the proof, we refer the reader to the arxiv version of this paper [2].

Proof of Theorem 1: We first observe that if we can find matrices A and B satisfying the assumptions of the theorem and sets S_1, S_2, Δ with $S_1 \subset S_2$ and $\Delta \subset S_2^c$ such that $\mathcal{E}_{\text{ave}}(A, B(S_1), +\infty) - \mathcal{E}_{\text{ave}}(A, B(S_1 \cup \Delta))$ is less than $\mathcal{E}_{\text{ave}}(A, B(S_2), +\infty) - \mathcal{E}_{\text{ave}}(A, B(S_2 \cup \Delta), +\infty)$, then we will have shown that $\mathcal{E}_{\text{ave}}(A, B(S), +\infty)$ is not a supermodular function of S. Indeed, this is almost identical to the definition of supermodularity with the inequality reversed, with the exception that the set Δ can now have more than a single element. However, if $\mathcal{E}_{\text{ave}}(A, B(S), +\infty)$ were supermodular, we could add the elements of Δ one by one to S_1 and S_2, respectively, and obtain that the right-hand side is at most the left-hand side in the above inequality.

We next describe how to choose A, B, S_1, S_2, Δ such that the above inequality holds. We mention again that choices will appear somewhat arbitrary; however, after the proof is over we will explain the intuition behind them.

The matrix B will be 2×5 and the matrix A will be 2×2. Furthermore, let us adopt the notation $b_1, \ldots, b_5$ for the five columns of B; each b_i belongs to $\mathbb{R}^2$.

First, we will set $A = (-1/2)I_2$. Observe that as a consequence of this

$$\mathcal{E}_{\text{ave}}(A, B, +\infty, S) = \sum_{i \in S} b_i b_i'.$$

Now, the columns of B will be determined as follows. Letting

$$W_{\text{init}} = \begin{pmatrix} 2^8 & 0 \\ 0 & 3 \cdot 2^9 \end{pmatrix}$$

and defining b_1, b_2 to be the vectors with the property that

$$b_1 b_1' + b_2 b_2' = W_{\text{init}}$$

specifically $b_1 = 2^4 e_1, b_2 = \sqrt{3 \cdot 2^9} e_2$. Similarly, let

$$W_\Delta = \begin{pmatrix} 5 \cdot 2^9 & -3 \cdot 2^9 \\ -3 \cdot 2^9 & 2^{10} \end{pmatrix}$$

and let b_3, b_4 be vectors such that

$$b_3 b_3' + b_4 b_4' = W_\Delta.$$

Such vectors exist because W_Δ is positive definite (this can be verified by looking at its two principal minors). Finally, we set $b_5 = [1\ 2^6]'$.

We now claim that

$$\mathcal{E}_{\text{ave}}(\{1,2\}) - \mathcal{E}_{\text{ave}}(\{1,2,3,4\}) < \mathcal{E}_{\text{ave}}(\{1,2,5\}) - \mathcal{E}_{\text{ave}}(\{1,2,3,4,5\}) \tag{5}$$

where $\mathcal{E}_{\text{ave}}(S)$ is used as shorthand for $\mathcal{E}_{\text{ave}}(A, B, +\infty, S)$ for the choices of A, B described above.

Indeed, since all the matrices are 2×2, we can compute both sides exactly. Using the identity

$$\text{tr} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{a+d}{ad-bc} \tag{6}$$

we compute expressions for the left- and right-hand sides of (4) in (6) and (7). A bit of calculation reveals that number on the right-hand side of (6) equals $49/14,208$, whereas the number of the right-hand side of (7), shown at the bottom of the page, equals $82,017,217/23,373,975,296$, and that the second of these numbers is bigger than the first. Thus, (5) follows

We next verify the conditions of the theorem. The matrix A is clearly strictly stable; unfortunately, it is not true that $W(A, B(S), +\infty)$ is always invertible.

To fix this define

$$A_\epsilon = -\frac{1}{2} I_2 + \epsilon 1_2 1_2'$$

where, recall, 1_2 is the vector of all-ones in $\mathbb{R}^2$. When ϵ is positive but small enough, we have that A_ϵ is strictly stable; furthermore we argue that when ϵ is small enough, $W(A, B(S)), +\infty)$ is then invertible for all nonempty S. Indeed, by the standard arguments, it suffices to show that the controllability matrix $[B(S)\ A_\epsilon B(S)]$ is invertible for all sets S, which contain only a single element. Now, since A_ϵ is 2×2, the only way the matrix $[b\ A_\epsilon b]$ could fail to be invertible is if $b = 0$ or b was an eigenvector of A_ϵ. Observe that the eigenvectors of A_ϵ are always $[1, 1]', [1, -1]'$, both of which we argue were avoided in our choice of the columns of B. Indeed, clearly the first, second, and fifth columns of B are clearly not proportional to either

$$\text{tr}\left[\begin{pmatrix} 2^8 & 0 \\ 0 & 3 \cdot 2^9 \end{pmatrix}^{-1} - \left(\begin{pmatrix} 2^8 & 0 \\ 0 & 3 \cdot 2^9 \end{pmatrix} + \begin{pmatrix} 5 \cdot 2^9 & -3 \cdot 2^9 \\ -3 \cdot 2^9 & 2^{10} \end{pmatrix}\right)^{-1}\right] = \frac{7}{2^9 \cdot 3} - \frac{3 \cdot 7}{2^9 \cdot 37} \tag{7}$$

$$\text{tr}\left[\left(\begin{pmatrix} 2^8 & 0 \\ 0 & 3 \cdot 2^9 \end{pmatrix} + \begin{pmatrix} 2^0 & 2^6 \\ 2^6 & 2^{12} \end{pmatrix}\right)^{-1} - \left(\begin{pmatrix} 2^8 & 0 \\ 0 & 3 \cdot 2^9 \end{pmatrix} + \begin{pmatrix} 2^0 & 2^6 \\ 2^6 & 2^{12} \end{pmatrix} + \begin{pmatrix} 5 \cdot 2^9 & -3 \cdot 2^9 \\ -3 \cdot 2^9 & 2^{10} \end{pmatrix}\right)^{-1}\right]$$
$$= \frac{3 \cdot 13 \cdot 151}{2^9 \cdot 2819} - \frac{9473}{2^9 \cdot 7^2 \cdot 661}. \tag{8}$$

of $[1,\ 1]', [1,\ -1]'$. As for the third and fourth columns, these were defined through the property that $b_3 b_3' + b_4 b_4' = W_\Delta$, so they can be chosen to be proportional to the eigenvectors of W_Δ, and it is easy to verify that neither $[1,\ 1]'$ nor $[1,\ -1]'$ is an eigenvector of W_Δ.

Finally, since $W(A, B(S), +\infty)$ is a continuous function of the entries of A over the set of strictly stable matrices,[2] we have that a counterexample may be picked by choosing ϵ small enough. ■

B. (Non)Supermodularity for Direct Variable Actuation

We now turn to the special case when B is the identity matrix. As we have previously remarked, this case has a special significance as it corresponds to choosing which variables can be directly actuated with an input.

Before stating our result, we introduce the following convention. Suppose that f is a function from $2^{\{1,\dots,n\}}$ to $\mathbb{R} \cup \{+\infty\}$. We will say that f is supermodular if (3) holds for all choices of $S_1 \subset S_2, a \in S_2^c$ such that every term in (3) is finite.

We now have the following theorem.

Theorem 2: There exists a strictly stable, symmetric matrix $A \in R^{6\times 6}$ such that $W(A, I_6(S), +\infty)$ is not a supermodular function of S.

Recall here our notation: I_6 refers to the 6×6 identity matrix and $I_6(S)$ is the matrix in $\mathbb{R}^{6\times|S|}$ obtained by picking the columns corresponding to the set $S \subset \{1, \dots, 6\}$.

Theorem 2 contradicts [5, Proposition 2]. Indeed, [5, Proposition 2] asserts that the function $\mathrm{tr}[(W(A, I(S), t) + \epsilon I)^{-1}]$ is supermodular, for ϵ small enough and any t. Taking the limit first as $t \to \infty$ and then as $\epsilon \to 0$, we obtain a contradiction with Theorem 2.

We next prove Theorem 2 by showing how the counterexample of Theorem 1 can be embedded into six dimensions.

Proof of Theorem 2: Our first observation is that the change of variables $y = Px$ does not change the control energy as long as P is orthonormal. Consequently, it suffices to construct a linear system with an orthonormal input matrix such that $W(\cdot, \cdot, , +\infty)$ is not supermodular, and then Theorem 2 will follow via a change of variables.

Take the matrix B constructed in that proposition. It is a 2×5 matrix; add one element to each row such that the two rows are orthogonal and have identical norm.[3] After this is done, normalize both rows to have unit norm. We now have a 2×6 counterexample whose rows are orthonormal. Call the resulting matrix B_1.

Define $A_1(K) = \mathrm{diag}(-K/2, -K/2, -4, -3, -2, -1)$. Let B_2 be the 6×6 matrix whose first two rows equal B_1 and the rest of the rows are equal to zero. Finally, we create B_3 by filling in random standard normal entries for the last four rows of B_2 and applying Gram–Schmidt to them. With probability one, we will thus have an orthonormal matrix whose first two rows are the same as the rows of B_1.

The motivation for this construction is as follows. We will later choose K to be very large, so that only what happens in the first two coordinates "matters" and the supermodularity of the system reduces to the supermodularity of the system in the first two components (which we already know does not hold by Theorem 1).

Let us adopt the notation that for a matrix M, we will use $\widehat{M}$ to denote its upper left 2×2 submatrix. Observe that, by construction, we have for any K that

$$\frac{K\widehat{W}(A_1(K), B_3(S), +\infty)}{W(A, B(S), +\infty)} = \text{ constant} \tag{9}$$

where the matrices A and B are taken from Theorem 1. Note that division of matrices is here understood elementwise. The constant on the right-hand side arises from the fact that the first two rows of B were normalized to obtain B_1.

We now argue that, with probability one, when K is large enough, we obtain the counterexample we seek in the pair $A_1(K)$ and B_3. The key step is the identity

$$\mathrm{tr}\begin{pmatrix} U & V \\ X & Y \end{pmatrix}^{-1} = \mathrm{tr}(U^{-1}) + \mathrm{tr}\left((Y - XU^{-1}V)^{-1}\right. \\ \left.(I + XU^{-2}V)\right) \tag{10}$$

which holds as long as U is invertible and $Y - XU^{-1}V$ is invertible [4]. Now, for any set S, let us partition the matrix $W(A_1, B_4(S), +\infty)$ as $\left(\begin{smallmatrix} U_W & X_W \\ V_W & Y_W \end{smallmatrix}\right)$ where its top 2×2 block is U_W.

First observe that, by (8), shown at the bottom of the previous page, for any $K > 0$ the matrix U_W is invertible as long as S is any of the sets in the counterexample of Theorem 1 (i.e., $S = \{1, 2\}, S = \{1, 2, 3, 4\}, S = \{1, 2, 5\}, S = \{1, 2, 3, 4, 5\}$), since the corresponding 2×2 matrices were computed to be invertible in the course of the proof of that theorem.

Moreover, as $K \to +\infty$, every nonzero entry of U_W, V_W, X_W goes to zero proportionately to $1/K$, whereas every entry of Y_W is constant. Thus, the matrix $Y_W - X_W U_W^{-1} V_W$ approaches Y_W. Since Y_W is invertible with probability 1 (this can be argued by first observing that it suffices to prove this when S is a singleton; and in that case, it follows from the observation that Y_W is a square submatrix of Hilbert matrix[4] scaled from the left and right by a random diagonal matrix whose entries have a zero probability of equaling zero), we obtain that with probability one, $Y_W - X_W U_W^{-1} V_W$ is invertible when K is large enough.

Consequently, on the right-hand side of (9), the second term is asymptotically negligible compared to the first one and we

[2]This follows because for strictly stable A, $W(A, B(S), +\infty)$ is the unique solution of the linear system equations $AW + WA' + BB' = 0$.

[3]This is always possible, since, if the two elements to be added (one to each row) are denoted as α and β, then they must satisfy $\alpha\beta = c_1, \alpha^2 - \beta^2 = c_2$, where c_1 is the negative inner product of the first two rows of B, and c_2 is the difference in the squared norm of the first two rows. Since the function $\alpha^2 - (c_1/\alpha)^2$ contains all of $\mathbb{R}$ in its range if $c_1 \neq 0$, such α and β can always be found. Finally, it is immediate to verify that indeed $c_1 \neq 0$ (i.e., the first two rows of the matrix B from Theorem 1 are not orthogonal).

[4]The Hilbert matrix is the matrix H defined by $H_{ij} = 1/(i + j - 1)$. It is known to be invertible, and indeed an explicit expression for its inverse is available; see for example http://mathworld.wolfram.com/HilbertMatrix.html.

obtain

$$\lim_{K\to\infty} \frac{\mathrm{tr}\begin{pmatrix} U_W & V_W \\ X_W & Y_W \end{pmatrix}^{-1}}{\mathrm{tr}(U_W^{-1})} = 1$$

Thus, as we choose K large enough, the average control energy of the system $\dot{x} = A_1(K)x + B_3(S)u$ will approach, in ratio, $\mathrm{tr}\, U_W^{-1}$, which is the same as $\mathrm{tr}[\widehat{W}(A_1, B_3(S), +\infty)^{-1}]$. Now applying (8), we see that the ratio of the average control energy of $\dot{x} = A_1(K)x + B_3(S)u$ to $K\mathrm{tr}(W(A, B(S), +\infty)^{-1}$ approaches a constant as $K \to +\infty$ for any of the sets S used in the proof of Theorem 1.

In other words, letting c denote the constant of the previous paragraph, we have that as $K \to +\infty$, the average control energy of $\dot{x} = A_1(K) + B_3(S)$ is $cK\mathrm{tr}(W(A, B(S), +\infty)^{-1}(1 + o_K(1))$, where $o_K(1)$ denotes something that approaches zero as $K \to +\infty$. Recall that here A, B are the matrices from Theorem 1.

We have already shown, however, the lack of supermodularity for $\mathrm{tr}(W(A, B(S), +\infty)^{-1}$ for these sets in Theorem 1, and thus we conclude that choosing K large enough we can obtain a counterexample to the average control energy $W(A_1, B_3(S), +\infty)$ using the same sets. ■

Remark: The matrix B constructed in this example is not uniquely defined, since it relies on the generation of random numbers. However, one run in MATLAB using the "randn" command to generate random Gaussians, with the choice of $K = 10^4$ yields (after rounding)

$$A = \begin{pmatrix} -182 & 0 & -565 & 0 & -11 & -736 \\ 0 & -1075 & 831 & -276 & -1752 & -612 \\ -565 & 831 & -2435 & 214 & 1321 & -1853 \\ 0 & -276 & 214 & -73 & -453 & -158 \\ -11 & -1752 & 1321 & -453 & -2864 & -1045 \\ -736 & -612 & -1853 & -158 & -1045 & -3381 \end{pmatrix}$$

with, of course, B being the 6×6 identity matrix. Using the MATLAB "gram" command to compute controllability Gramians, we obtain

$$\mathcal{E}_{\mathrm{ave}}(\{1,2\}) - \mathcal{E}_{\mathrm{ave}}(\{1,2,3,4\}) \approx 2.50 \cdot 10^5$$
$$\mathcal{E}_{\mathrm{ave}}(\{1,2,5\}) - \mathcal{E}_{\mathrm{ave}}(\{1,2,3,4,5\}) \approx 2.52 \cdot 10^5$$

providing a numerical confirmation of nonsupermodularity for this example.

IV. Conclusion

We have constructed two examples showing that average control energy is not necessarily a supermodular function of the set of actuated sites or actuated variables. These results are relevant for the problem of actuator placement with average energy constraints; in that they show that a key property that has been used to develop approximation algorithms in other contexts is not available here.

Indeed in [1], it was shown that if actuating the variables in the set S^* renders a system controllable, then one can find in polynomial time a set of size $O(|S^*| \log n)$ that also renders the system controllable, and moreover this is the best possible guarantee one can obtain in polynomial time unless $P = NP$. The proof was based on the submodularity of the dimension of the controllable subspace. It is at present unclear what the analogous best possible guarantee one can attain (in polynomial time) when the control metric is not controllability of the system but rather average control energy.

References

[1] A. Olshevsky, "Minimal controllability problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 249–258, Sep. 2014.
[2] A. Olshevsky, On(Non)supermodularity of average control energy. 2016. [Online]. Available: https://arxiv.org/abs/1609.08706
[3] T. Summers, F. 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.
[4] T. Tao, Matrix identities as derivatives of determinant identities. 2013. [Online]. Available: http://tinyurl.com/h7umlga
[5] 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.

Alex Olshevsky received the B.S. degrees in applied mathematics and electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, both in 2004, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2006 and 2010, respectively.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA. His research interests include control systems, optimization, and network science.

Dr. Olshevsky received the National Science Foundation CAREER Award, the Air Force Young Investigator Award, the ICS Prize from INFORMS for best paper on the interface of operations research and computer science, and the SIAM paper prize for annual paper from the *SIAM Journal on Control and Optimization* chosen to be reprinted in *SIAM Review*.

Minimal Reachability is Hard To Approximate

Ali Jadbabaie, Alexander Olshevsky, George J. Pappas, and Vasileios Tzoumas

***Abstract*—In this note, we consider the problem of choosing, which nodes of a linear dynamical system should be actuated so that the state transfer from the system's initial condition to a given final state is possible. Assuming a standard complexity hypothesis, we show that this problem cannot be efficiently solved or approximated in polynomial, or even quasi-polynomial, time.**

***Index Terms*—Approximation algorithms, computational complexity, controllability, (non-)submodularity, sparse actuator placement.**

I. Introduction

During the last decade, researchers in systems, optimization, and control have focused on the questions.

1) *Actuator Selection:* How many nodes do we need to actuate in a gene regulatory network to control it? [1], [2]
2) *Input Selection:* How many inputs are needed to drive the nodes of a power system to fully control its dynamics? [3]
3) *Leader Selection:* Which UAVs do we need to choose in a multi-UAV system as leaders for the system to complete a surveillance task despite communication noise? [4], [5]

The effort to answer such questions has resulted in numerous papers on topics such as actuator placement for controllability [6], [7]; actuator selection and scheduling for bounded control effort [8]–[11]; resilient actuator placement against failures and attacks [12], [13]; and sensor selection for target tracking and optimal Kalman filtering [14]–[17]. In all these papers, the underlying optimization problems have been proven (i) either polynomially-time solvable [1]–[3] (ii) or NP-hard, in which case polynomial-time algorithms have been proposed for their approximate solution [4]–[17].

But in systems, optimization, and control, such as in power systems [18], [19], transportation networks [20], and neural circuits [21], [22], the following problem also arises:

Minimal reachability problem: Given times t_0 and t_1 such that $t_1 > t_0$, vectors x_0 and x_1, and a linear dynamical system with state vector $x(t)$ such that $x(t_0) = x_0$, find the minimal number of system nodes we need to actuate so that the state transfer from $x(t_0) = x_0$ to $x(t_1) = x_1$ is feasible.

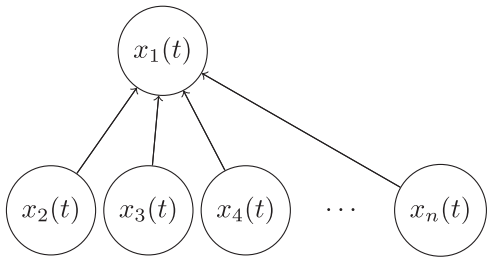


Fig. 1. Graphical representation of the linear system $\dot{x}_1(t) = \sum_{j=2}^{n} x_j(t)$, $\dot{x}_i(t) = 0$, $i = 2, \ldots, n$; each node represents an entry of the system's state $(x_1(t), x_2(t), \ldots, x_n(t))$, where t represents time; the edges denote that the evolution in time of x_1 depends on $(x_2, x_3, \ldots, x_n)$.

For example, the stability of power systems is ensured by placing a few generators such that the state transfers from a set of possible initial conditions to the zero state are feasible [19].

The minimal reachability problem relaxes the objectives of the applications in [1]–[17]. For example, in comparison to the actuator placement problem for controllability [6], the minimal reachability problem aims to place a few actuators only to make a single transfer between two states feasible, whereas the minimal controllability problem aims to place a few actuators to make the transfer among any two states feasible [6], [7].

The fact that the minimal reachability problem relaxes the objectives of the note [1]–[17] is an important distinction whenever we are interested in the feasibility of only a few state transfers by a small number of placed actuators. The reason is that under the objective of minimal reachability the number of placed actuators can be much smaller in comparison to the number of placed actuators under the objective of controllability. For example, in the system of Fig. 1 the number of placed actuators under the objective of minimal reachability from $(0, \ldots, 0)$ to $(1, \ldots, 0)$ is one, whereas the number of placed actuators under the objective of controllability grows linearly with the system's size.

The minimal reachability problem was introduced in [23], where it was found to be NP-hard. Similar versions of the reachability problem were studied in the context of power systems in [19] and [24]. For the polynomial-time solution of the reachability problems in [19], [23], [24], greedy approximation algorithms were proposed therein. The approximation performance of these algorithms was claimed by relying on the modularity result [25, Lemma 8.1], which states that the distance from a point to a subspace created by the span of a set of vectors is supermodular in the choice of the vectors.

In this note, we first show that the modularity result [25, Lemma 8.1] is incorrect. In particular, we show this via a counterexample to [25, Lemma 8.1], and as a result, we prove that the distance from a point to a subspace created by the span of a set of vectors is nonsupermodular

Manuscript received October 29, 2017; revised October 30, 2017 and February 20, 2018; accepted May 3, 2018. Date of publication May 14, 2018; date of current version January 28, 2019. This work was supported in part by the Vannevar Bush Fellowship from Office of Secretary of State, in part by the National Science Foundation under Grant ECCS-1740451 and the Army Research Office W911NF-18-1-0072. Recommended by Associate Editor S. Miani. *(Corresponding author: Vasileios Tzoumas.)*

A. Jadbabaie is with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge MA 02139 USA (e-mail: jadbabai@mit.edu).

A. Olshevsky is with the Department of Electrical and Computer Engineering and the Division of Systems Engineering, Boston University, Boston, MA 02215 USA (e-mail: alexols@bu.edu).

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

Digital Object Identifier 10.1109/TAC.2018.2836021

in the choice of the vectors. Then, we also prove the following strong intractability result for the minimal reachability problem, which is our main contribution in this note.

Contribution 1: Assuming NP $\notin$ BPTIME$(n^{\text{poly}\log n})$, we show that for each $\delta > 0$, there is no polynomial-time algorithm that can distinguish[1] between the two cases in which the following conditions hold.

1) The reachability problem has a solution with cardinality k.
2) The reachability problem has no solution with cardinality $k2^{\Omega\left(\log^{1-\delta} n\right)}$, where n is the dimension of the system.

We note that the complexity hypothesis NP $\notin$ BPTIME $(n^{\text{poly}\log n})$ means there is no randomized algorithm which, after running for $O(n^{(\log n)^c})$ time for some constant c, outputs correct solutions to problems in NP with probability $2/3$; see [26] for more details.

Notably, Contribution 1 remains true even if we allow the algorithm to search for an approximate solution that is relaxed as follows: Instead of choosing the actuators to make the state transfer from the initial state x_0 to a given final state x_1 possible, some other state $\widehat{x}_1$ that satisfies $\|x_1 - \widehat{x}_1\|_2^2 \leq \epsilon$ should be reachable from x_0. This is a substantial relaxation of the reachability problem's objective, and yet, we show that the intractability result of Contribution 1 still holds.

The rest of this note is organized as follows. In Section II, we introduce formally the minimal reachability problem. In Section III, we provide a counterexample to [25, Lemma 8.1]. In Section IV, we present Contribution 1; in Section V, we prove it. Section VI concludes the note.

II. Minimal Reachability Problem

In this section, we formalize the minimal reachability problem. We start by introducing the systems considered in this note and the notions of system node and of actuated node set.

System: We consider linear systems of the form

$$\dot{x}(t) = Ax(t) + Bu(t),\ t \geq t_0 \tag{1}$$

where t_0 is a given starting time, $x(t) \in \mathbb{R}^n$ is the system's state at time t, and $u(t) \in \mathbb{R}^m$ is the system's input vector.

In this note, we want to actuate the minimal number of the nodes of the system in (1) to make a desired state-transfer feasible (not achieving necessarily controllability). We formalize this objective using the following two definitions.

Definition 1 (System node): Given a system as in (1), where $x(t) \in \mathbb{R}^n$, we let $x_1(t), x_2(t), \ldots, x_n(t) \in \mathbb{R}$ be the components of $x(t)$, i.e., $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))$. We refer to each $x_i(t)$ as a *system node*.

Definition 2 (Actuated node set): Given a system as in (1), we say that the set $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$ is an *actuated node set* if the system dynamics can be written as

$$\dot{x}(t) = Ax(t) + \mathbb{I}(\mathcal{S})Bu(t),\ t \geq t_0 \tag{2}$$

where $\mathbb{I}(\mathcal{S})$ is a diagonal matrix such that if $i \in \mathcal{S}$, the ith entry of $\mathbb{I}(\mathcal{S})$'s diagonal is 1, otherwise it is 0.

The definition of $\mathbb{I}(\mathcal{S})$ in (2) implies that the input $u(t)$ affects only the system nodes $x_i(t)$ where $i \in \mathcal{S}$.

[1]We say that an algorithm can distinguish between two (disjoint) cases A and B if, when fed with an input that is guaranteed to be in either A or B, the algorithm is able to determine which of the two is the case (e.g., by outputing 1 if the input belongs A, and 0 if it belongs to B).

Problem 1 (Minimal Reachability): Given

1) times t_0 and t_1 such that $t_1 > t_0$,
2) vectors $x_0,\ x_1 \in \mathbb{R}^n$, and
3) a system $\dot{x}(t) = Ax(t) + Bu(t)$, $t \geq t_0$, as in (1), with initial condition $x(t_0) = x_0$,

find an actuated node set with minimal cardinality such that there exists an input $u(t)$ defined over the time interval (t_0, t_1) that achieves $x(t_1) = x_1$. Formally, using the notation $|\mathcal{S}|$ to denote the cardinality of a set $\mathcal{S}$

$$\underset{\mathcal{S}\subseteq\{1,2,\ldots,n\}}{\text{minimize}} \quad |\mathcal{S}|$$

such that there exist $u : (t_0, t_1) \mapsto \mathbb{R}^m$, $x : (t_0, t_1) \mapsto \mathbb{R}^n$ with

$$\begin{aligned} \dot{x}(t) &= Ax(t) + \mathbb{I}(\mathcal{S})Bu(t),\ t \geq t_0 \\ x(t_0) &= x_0,\ x(t_1) = x_1. \end{aligned}$$

A special case of interest is when B is the identity matrix. Then, minimal reachability asks for the fewest system nodes to be actuated directly so that at time t_1 the state x_1 is reachable from the system's initial condition $x(t_0) = x_0$.

III. Nonsmodularity of Distance From Point to Subspace

In this section, we provide a counterexample to the supermodularity result [25, Lemma 8.1]. We begin with some notation. In particular, given a matrix $M \in \mathbb{R}^{n\times n}$, a vector $v \in \mathbb{R}^n$, and a set $\mathcal{S} \subset \{1, \ldots, n\}$, let $M(\mathcal{S})$ denote the matrix obtained by throwing away columns of M not in $\mathcal{S}$. In addition, for any set $\mathcal{S} \subset \{1, \ldots, n\}$, we define the set function

$$f(\mathcal{S}) = \text{dist}^2(v, \text{Range}(M(\mathcal{S})))$$

where $\text{dist}(y, X)$ is the distance from a point to a subspace

$$\text{dist}(y, X) = \min_{x\in X} ||y - x||_2.$$

We show there exist a vector v and a matrix M such that the function $f : 2^{\{1,2,\ldots,n\}} \mapsto \text{dist}^2(v, \text{Range}(M(\mathcal{S})))$ is nonsupermodular. We start by defining the monotonicity and supermodularity of set functions.

Definition 3 (Monotonicity): Consider any finite set $\mathcal{V}$. The set function $f : 2^{\mathcal{V}} \mapsto \mathbb{R}$ is nondecreasing if and only if for any $\mathcal{A} \subseteq \mathcal{A}' \subseteq \mathcal{V}$, we have $f(\mathcal{A}) \leq f(\mathcal{A}')$.

In words, a set function $f : 2^{\mathcal{V}} \mapsto \mathbb{R}$ is nondecreasing if and only if by adding elements in any set $\mathcal{A} \subseteq \mathcal{V}$ we cannot decrease the value of $f(\mathcal{A})$.

Definition 4 (Supermodularity [27, Proposition 2.1]): Consider any finite set $\mathcal{V}$. The set function $f : 2^{\mathcal{V}} \mapsto \mathbb{R}$ is supermodular if and only if for any $\mathcal{A} \subseteq \mathcal{A}' \subseteq \mathcal{V}$ and $x \in \mathcal{V}$

$$f(\mathcal{A}) - f(\mathcal{A} \cup \{x\}) \geq f(\mathcal{A}') - f(\mathcal{A}' \cup \{x\}).$$

In words, a set function $f : 2^{\mathcal{V}} \mapsto \mathbb{R}$ is supermodular if and only if it satisfies the following diminishing returns property: for any element $x \in \mathcal{V}$, the marginal decrease $f(\mathcal{A}) - f(\mathcal{A} \cup \{x\})$ diminishes as the set $\mathcal{A}$ grows; equivalently, for any $\mathcal{A} \subseteq \mathcal{V}$ and $x \in \mathcal{V}$, $f(\mathcal{A}) - f(\mathcal{A} \cup \{x\})$ is nonincreasing.

Example 1: We show that for

$$v = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix},\ M = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

the set function $f : 2^{\{1,2,3\}} \mapsto \text{dist}^2(v, \text{Range}(M(\mathcal{S})))$ is nonsupermodular. In particular, since the vector v is orthogonal to the first and

third columns of M,

$$f(\{1\}) = \text{dist}^2(v, M(\{1\})) = ||v||_2^2$$
$$f(\{1,3\}) = \text{dist}^2(v, M(\{1,3\})) = ||v||_2^2$$

Therefore,

$$f(\{1\}) - f(\{1,3\}) = 0.$$

At the same time, the span of the first two columns of M is the subspace $\{x \in \mathbb{R}^3 : x_3 = 0\}$. Thus,

$$f(\{1,2\}) = \text{dist}^2(v, M(\{1,2\})) = 1.$$

Also, since the three columns of M are linearly independent

$$f(\{1,2,3\}) = \text{dist}^2(v, M(\{1,2,3\})) = 0$$

and as a result

$$f(\{1,2\}) - f(\{1,2,3\}) = 1.$$

In sum

$$f(\{1,2\}) - f(\{1,2,3\}) > f(\{1\}) - f(\{1,3\}).$$

Hence, for the vector v and matrix M in this example, $f : 2^{\{1,2,3\}} \mapsto \text{dist}^2(v, \text{Range}(M(\mathcal{S})))$ is nonsupermodular.

We remark that the same argument as in Example 1 shows that the set function $g : 2^{\{1,2,\ldots,n\}} \mapsto \mathbb{R}$ such that $g(\mathcal{S}) = [\text{dist}(v, \text{Range}(M(\mathcal{S}))]^c$ is not supermodular for any $c > 0$. It is also possible to see that $g(\mathcal{S})$ is not submodular: e.g., consider the case where M repeats the same columns.

IV. Inapproximability of Minimal Reachability Problem

We show that, subject to a widely believed conjecture in complexity theory, there is no efficient algorithm that solves, even approximately, Problem 1. Towards the statement of this result, we next introduce a definition of approximability and the definition of quasi-polynomial running time.

Definition 5 (Approximability): Consider the minimal reachability Problem 1, and let the set $\mathcal{S}^\star$ to denote one of its optimal solutions. We say that an algorithm renders Problem 1 $(\Delta_1(n), \Delta_2(n))$-*approximable* if it returns a set $\mathcal{S}$ such that.

1) There is a state $\widehat{x}_1$ such that there is an input $u(t)$ such that at time t_1 we have $x(t_1) = \widehat{x}_1$ and $||\widehat{x}_1 - x_1||_2 \leq \Delta_1(n)$.
2) The cardinality of the set $\mathcal{S}$ is at most $\Delta_2(n)|\mathcal{S}^\star|$.

Hence, the definition of $(\Delta_1(n), \Delta_2(n))$-approximability allows some slack both in the quality of the reachability requirement (first point in the itemization in Definition 5), and in the number of actuators utilized to achieve it (second point in the itemization in Definition 5).

We introduce next the definition of quasi-polynomial algorithms, using the following big O notation.

Definition 6 (Big O notation): Let $\mathbb{N}$ be the set of natural numbers, and consider two functions $h : \mathbb{N} \mapsto \mathbb{R}$ and $g : \mathbb{N} \mapsto \mathbb{R}$ that take only nonnegative values. The *big O notation* in the equality $h(n) = O(g(n))$ means there exists some constant $c > 0$ such that for all large enough n, it is $h(n) \leq cg(n)$.

Definition 6, given a nonnegative function g, implies that $O(g(n))$ denotes the collection of nonnegative functions h that are bounded asymptotically by g, up to a constant factor.

Definition 7 (Quasi-polynomial running time): An algorithm is *quasi-polynomial* if it runs in $2^{O[(\log n)^c]}$ time, where c is a constant.

We note that any polynomial-time algorithm is a quasi-polynomial time algorithm since $n^k = 2^{k \log n}$. At the same time, a quasi-polynomial algorithm is asymptotically faster than an exponential-time algorithm, since exponential-time algorithms run in $O(2^{n^\epsilon})$ time, for some $\epsilon > 0$.

Definition 8 (Big Omega notation): Let $\mathbb{N}$ be the set of natural numbers, and consider the functions $h : \mathbb{N} \mapsto \mathbb{R}$ and $g : \mathbb{N} \mapsto \mathbb{R}$ that take only nonnegative values. The *big Omega notation* in the equality $h(n) = \Omega(g(n))$ means that there exists some constant $c > 0$ such that for all large enough n, it is $h(n) \geq cg(n)$.

Definition 8, given a nonnegative g, implies that $\Omega(g(n))$ denotes the collection of nonnegative functions h that are lower bounded asymptotically by g, up to a constant factor.

We present next our main result in this note.

Theorem 1 (Inapproximability): For each $\delta \in (0,1)$, there is a collection of instances of Problem 1 where the following conditions hold.

1) The initial condition is $x(t_0) = 0$.
2) The final state x_1 is of the form $[1, 1, \ldots, 1, 0, 0, \ldots, 0]^\top$.
3) The input matrix is $B = I$, where I is the identity matrix

along with a polynomial $\Delta_1(n)$ and a function $\Delta_2(n) = 2^{\Omega(\log^{1-\delta} n)}$, such that unless NP$\in$ BPTIME$(n^{\text{poly}\log n})$, there is no quasi-polynomial algorithm rendering Problem 1 $(\Delta_1(n), \Delta_2(n))$-approximable.

Theorem 1 says that if NP $\notin$ BPTIME$(n^{\text{poly}\log n})$ there is no polynomial time algorithm (or quasi-polynomial time algorithm) that can choose which entries of the system's x state to actuate so that $x(t_1)$ is even approximately close to a desired state $x_1 = [1, 1, \ldots, 1, 0, 0, \ldots, 0]^\top$ at time t_1.

To make sense of Theorem 1, first observe that we can always actuate every entry of the system's state, i.e., we can choose $\mathcal{S} = \{1, 2, \ldots, n\}$. This means every system is $(0, n)$-approximable; let us rephrase this by saying that every system is $(0, 2^{\log n})$ approximate. Theorem 1 tells us that we cannot achieve $(0, 2^{O(\log^{1-\delta} n)})$-approximability *for any* $\delta > 0$. In other words, improving the guarantee of the strategy that actuates every state by just a little bit, in the sense of replacing $\delta = 0$ with some $\delta > 0$, is not possible —subject to the complexity-theoretic hypothesis NP $\notin$ BPTIME$(n^{\text{poly}\log n})$. Furthermore, the theorem tells us it remains impossible even if we allow ourselves some error $\Delta(n)$ in the target state, i.e., even $(\Delta(n), 2^{O(\log^{1-\delta} n)})$–approximability is ruled out.

Remark 1: In [23, Theorem 3] it is claimed that for any $\epsilon > 0$ the minimal reachability Problem 1 is $\left(\epsilon, O\left(\log \frac{n}{\epsilon}\right)\right)$-approximable, which contradicts Theorem 1. However, the proof of this claim was based on [25, Lemma 8.1], which we proved incorrect in Section III.

Remark 2: The minimal controllability problem [6] seeks to place the fewest number of actuators to make the system controllable. Theorem 1 is arguably surprising, as it was shown in [6] that the sparsest set of actuators for controllability can be approximated to a multiplicative factor of $O(\log n)$ in polynomial time. By contrast, we showed in this note that an almost exponentially worse approximation ratio *cannot* be achieved for minimum reachability.

V. Proof of Inapproximability of Minimal Reachability

We next provide a proof of our main result, namely Theorem 1. We use some standard notation throughout: $\mathbf{1}_k$ is the all-ones vector in $\mathbb{R}^k$, $\mathbf{0}_k$ is the zero vector in $\mathbb{R}^k$, and e_k is the kth standard basis vector. We begin with some standard definitions related to the reachability space of a linear system.

A. Reachability Space for Continuous-Time Linear Systems

Definition 9 (Reachability space): Consider a system $\dot{x}(t) = Ax(t) + Bu(t)$ as in (1) whose size is n. The Range

$([B,\ AB,\ A^2B, \ldots,\ A^{n-1}B])$ is called the *reachability space* of $\dot{x}(t) = Ax(t) + Bu(t)$.

The reason why Definition 9 is called the reachability space is explained in the following proposition.

Proposition 1 ([28, Proof of Theorem 6.1]): Consider a system as in (1), with initial condition x_0. There exists a real input $u(t)$ defined over the time interval (t_0, t_1) such that the solution of $\dot{x} = Ax + Bu$, $x(t_0) = x_0$ satisfies $x(t_1) = x_1$ if and only if

$$x_1 - e^{A(t_1 - t_0)}x_0 \in \text{Range}([B,\ AB,\ A^2B, \ldots,\ A^{n-1}B]).$$

The notion of reachability space allows us to redefine the minimal reachability Problem 1 as follows.

Corollary 1: Problem 1 is equivalent to

$$\begin{aligned} &\underset{\mathcal{S}\subseteq\{1,2,\ldots,n\}}{\text{minimize}}\ |\mathcal{S}| \\ &\text{such that}\ \ x_1 - e^{A(t_1-t_0)}x_0 \in \\ &\qquad \text{Range}([\mathbb{I}(\mathcal{S})B,\ A\mathbb{I}(\mathcal{S})B, \ldots,\ A^{n-1}\mathbb{I}(\mathcal{S})B]). \end{aligned}$$

Overall, Problem 1 is equivalent to picking the fewest rows of the input matrix B such that $x_1 - e^{A(t_1-t_0)}x_0$ is in the linear span of the columns of $[\mathbb{I}(\mathcal{S})B,\ A\mathbb{I}(\mathcal{S})B,\ A^2\mathbb{I}(\mathcal{S})B, \ldots,\ A^{n-1}\mathbb{I}(\mathcal{S})B]$.

B. Variable Selection Problem

We show the intractability of the minimum reachability by reducing it to the *variable selection* problem, defined next.

Problem 2 (Variable Selection): Let $U \in \mathbb{R}^{m\times l}$, $z \in \mathbb{R}^m$, and let Δ be a positive number. The variable selection problem is to pick $y \in \mathbb{R}^l$ that is an optimal solution to the following optimization problem.

$$\begin{aligned} \underset{y\in\mathbb{R}^l}{\text{minimize}} \quad & \|y\|_0 \\ \text{such that} \quad & \|Uy - z\|_2 \le \Delta \end{aligned}$$

where $||y||_0$ refers to the number of nonzero entries of y.

The variable selection Problem 2 is found in [29] to be inapproximable, even in quasi-polynomial time.

Theorem 2 ([29, Proposition 6]): Unless it is NP$\in$ BPTIME$(n^{\text{poly}\log n})$, for each $\delta \in (0,1)$ there exist the following:

1) a function $q_1(l)$ which is in $2^{\Omega(\log^{1-\delta} l)}$;
2) a polynomial $p_1(l)$ which is in $O(l)$;[2]
3) a polynomial $\Delta(l)$;
4) a polynomial $m(l)$,

and a zero one $m(l) \times l$ matrix U such that no quasi-polynomial time algorithm can distinguish between the following two cases for large l.

1) There exists a vector $y \in \mathbb{R}^l$ such that $Uy = \mathbf{1}_{m(l)}$ and $||y||_0 \le p_1(l)$.
2) For any vector $y \in \mathbb{R}^l$ such that $||Uy - \mathbf{1}_{m(l)}||_2^2 \le \Delta(l)$, we have $||y||_0 \ge p_1(l)q_1(l)$.

Informally, unless NP$\in$BPTIME$(n^{\text{poly}\log n})$, Theorem 2 says that Problem 2 is inapproximable even in quasi-polynomial time, in the sense that for large l there is no quasi-polynomial algorithm that can distinguish between the two mutually exclusive cases 1) and 2). To see that these cases are indeed mutually exclusive for large l, observe that $q_1(l) > 1$ when l is large, because $q_1(l) = 2^{\Omega\left(\log^{1-\delta} l\right)}$.

[2]In this context, a function with a fractional exponent is considered to be a polynomial, e.g., $l^{1/5}$ is considered to be a polynomial in l.

C. Sketch of Proof of Theorem 1

We begin by sketching the intuition behind the proof of Theorem 1. Our general approach is to find instances of Problem 1 that are as hard as inapproximable instances of the variable selection Problem 2. We begin by discussing a construction that does *not* work, and then explain how to fix it.

Given the matrix U coming from a variable selection Problem 2, we first attempt to construct an instance of the minimal reachability Problem 1 where the following conditions hold.

1) The system's initial condition is $x(t_0) = 0$.
2) The destination state x_1 at time t_1 is of the form $[\mathbf{1}, \mathbf{0}]^\top$ (the exact dimensions of $\mathbf{1}$ and $\mathbf{0}$ are to be determined).
3) The input matrix is $B = I$.
4) The system matrix A is

$$A = \begin{pmatrix} 0 & U \\ 0 & 0 \end{pmatrix} \tag{3}$$

where the number of zeros is large so that $A^2 = 0$.

Whereas the variable selection problem involves finding the smallest set of columns of U so that a certain vector is in their span, for the minimum reachability problem, every time we add the kth state to the set of actuated variables $\mathcal{S}$, the reachability span expands by adding the span of the set of columns of the controllability matrix that correspond to the vector e_k being added in $\mathbb{I}(\mathcal{S})$. In particular, for the above construction, because $A^2 = 0$, when the kth state is added to the set of actuated variables, the span of the two columns e_k and Ue_k is added to the reachability space.

In other words, with the above construction we are basically constrained to make "moves," which add columns in pairs, and we are looking for the smallest number of such "moves" making a certain vector lie in the span of the columns. It should be clear that there is a strong parallel between this and variable selection (where the columns are added one at a time). However, because the columns are being added in pairs, this attempt to connect minimum reachability with variable selection does not quite work. To fix this idea, we want only the columns of U to contribute meaningfully to the addition of the span, with any vectors e_k we add along the way being redundant; this would reduce minimal reachability to exactly variable selection. We accomplish this by further defining

$$U' = \begin{pmatrix} U \\ U \\ \vdots \\ U \end{pmatrix}$$

where we stack U some large number of times (to be determined in the main proof of Theorem 1 at Section V-D). We then set

$$A = \begin{pmatrix} 0 & U' \\ 0 & 0 \end{pmatrix}. \tag{4}$$

The idea is because U is stacked many times, adding a column of U to a set of vectors expands the span much more than adding any vector e_k, so there is never an incentive to consider the contributions of any e_k to the reachability space.

We make the aforementioned construction of the system matrix A precise: given a matrix $M \in \mathbb{R}^{m\times l}$, for $n \ge \max\{m, l\}d$ we define $\phi_{n,d}(M)$ to be the $n \times n$ matrix which stacks M in the top right-hand

corner d times. For example

$$M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \ \phi_{5,2}(M) = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

i.e., $\phi_{5,2}(M)$ stacks M twice, and then pads it with enough zeros to make the resulting matrix 5×5. Observe that $\phi_{n,d}(M)^2 = 0$ for $n \geq \max\{m, l\}(d+1)$. Overall, in Section VI, we set $A = \phi_{n,d}(U)$ for large enough d, and $n = \max\{m, l\}(d+1)$, and we prove Theorem 1.

D. Proof of Theorem 1

Adopting the notation in Theorem 2, we focus on problem instances where for large enough l it is $q_1(l) > 1$, per the proof of Theorem 2, i.e., of [29, Theorem 2]. In addition, we let $d = \lceil p_1(l)q_1(l) \rceil$, and $n = \max\{m(l), l\}(d+1)$. Moreover, for simplicity, we use henceforth m and $m(l)$ interchangeably. Finally, we consider the instances of Problem 1 where the following conditions hold.

1) The initial condition is $x(t_0) = \mathbf{0}_n$.
2) The destination state x_1 at time t_1 is $[\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top$.
3) The input matrix is $B = I$, where I is the identity matrix.
4) The system matrix is $A = \phi_{n,d}(U)$.

Given the above-mentioned conditions, to prove Theorem 1 we first define the following four statements:

S1) There exists a vector $y \in \mathbb{R}^l$ such that $Uy = \mathbf{1}_m$ and $||y||_0 \leq p_1(l)$.
S2) For any vector $y \in \mathbb{R}^l$ such that $||Uy - \mathbf{1}_m||_2^2 \leq \Delta(l)$, we have $||y||_0 \geq p_1(l)q_1(l)$.
S1′) There exists a set $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$ with $|\mathcal{S}| \leq p_1(l)$ such that the state $x_1 = [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top$ is reachable.
S2′) There is no set $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$ with cardinality strictly less than $p_1(l)q_1(l)$ that makes reachable some $\widehat{x}_1$ with $||\widehat{x}_1 - [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top||_2^2 \leq \Delta(l)$.

Recall that in Section V-B, we stated that the statements S1-S2 are mutually exclusive for $q_1(l) > 1$ (which is the case for the instances we consider in this proof), and that Theorem 2 implies there is no quasi-polynomial algorithm (unless NP$\in$ BPTIME$(n^{\text{poly}\log n})$) that can distinguish between S1 and S2.

Given the above, we next proceed with the proof of Theorem 1 by proving first that statement S1 implies statement S1′, and then that also statement S2 implies statement S2′.

Proof that statement S1 implies statement S1′: We prove that if statement S1 is true, then statement S1′ also is. In particular, suppose there exists a vector $y \in \mathbb{R}^l$ with $Uy = \mathbf{1}_m$ and $||y||_0 \leq p_1(l)$ (statement S1). In this case, we claim there exists a set $\mathbb{S} \subseteq \{1, 2, \ldots, n\}$ with $|\mathbb{S}| \leq p_1(l)$ such that $x_1 = [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top$ is reachable (statement S1′). Indeed, let S be a set of columns of U that have $\mathbf{1}_m$ in their span, and set $\mathbb{S} = \{k + n - l \,|\, k \in S\}$. Then $|\mathbb{S}| \leq p_1(l)$, and

$$\mathbf{1}_m = \sum_{k \in S} y_k U_k \tag{5}$$

where y_k denotes the kth element of the vector y, and U_k denotes the kth column of the matrix U. Due to (5), we can rewrite the vector $x_1 = [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top$ as follows:

$$\begin{pmatrix} \mathbf{1}_{m\,d} \\ \mathbf{0}_{n-m\,d} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_m \\ \mathbf{1}_m \\ \vdots \\ \mathbf{1}_m \\ \mathbf{0}_{n-m\,d} \end{pmatrix} = \sum_{k \in S} \begin{pmatrix} y_k U_k \\ y_k U_k \\ \vdots \\ y_k U_k \\ \mathbf{0}_{n-m\,d} \end{pmatrix} = \sum_{k \in S} y_k A_{k+n-l}, \tag{6}$$

where the vector $\mathbf{1}_m$ in the second term from the left is repeated $\lceil p_1(l)q_1(l) \rceil$ times, since $d = \lceil p_1(l)q_1(l) \rceil$, and where the final step (6) follows by definitions of A as $A = \phi_{n,d}(U)$, and where A_{k+n-l} denotes the $(k+n-l)$th column of A. Now, each of the vectors A_{k+n-l} in the last term is a column of $A\mathbb{I}(\mathbb{S})$, so $[\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top$ indeed lies in the range of the controllability matrix and, as a result, the state $x_1 = [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top$ is reachable by actuating $\mathbb{S}$.

Proof that statement S2 implies statement S2′: We prove that if the statement S2 is true, then the statement S2′ also is. In particular, per statement S2 suppose that any vector y with $||Uy - \mathbf{1}_m||_2^2 \leq \Delta(l)$ has the property that $||y||_0 \geq p_1(l)q_1(l)$. We claim that in this case there is no set $\mathbb{S} \subseteq \{1, 2, \ldots, n\}$ with cardinality strictly less than $p_1(l)q_1(l)$ that makes reachable some $\widehat{x}_1$ with $||\widehat{x}_1 - [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top||_2^2 \leq \Delta(l)$ (statement S2′). To prove this, assume the contrary, i.e., assume there exists $\mathbb{S}$ with cardinality strictly less than $p_1(l)q_1(l)$ that makes reachable some $\widehat{x}_1$ with $||\widehat{x}_1 - [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top||_2^2 \leq \Delta(l)$ —we call this assumption A1. We obtain a contradiction as follows: the pigeonhole principle implies that in the set $\{1, 2, \ldots, md\}$ there is some interval $\mathbb{E} = \{\kappa m + 1, \kappa m + 2, \ldots, \kappa m + m\}$, where κ is a non-negative integer, such that $\mathbb{S} \cap \mathbb{E} = \emptyset$, because $|\mathbb{S}| < p_1(l)q_1(l)$ and $md \geq m\lceil p_1(l)q_1(l) \rceil$. Define the vector $\widehat{x}_{\mathbb{E}} \in \mathbb{R}^m$ by taking the rows of $\widehat{x}_1$ corresponding to indexes in $\mathbb{E}$. Then,

$$||[\widehat{x}_{\mathbb{E}} - \mathbf{1}_m||_2^2 \leq \Delta(l)$$

since $\widehat{x}_1$ with $||\widehat{x}_1 - [\mathbf{1}_{m\,d}^\top, \mathbf{0}_{n-m\,d}^\top]^\top||_2^2 \leq \Delta(l)$. Moreover, we next prove that $\widehat{x}_{\mathbb{E}}$ is in the span of $|\mathbb{S}|$ columns of U. To this end, we make the following observations: since $\widehat{x}_1$ is reachable, it is

$$\begin{aligned} \widehat{x}_1 \in & \ \text{Range}[\mathbb{I}(\mathbb{S}), A\mathbb{I}(\mathbb{S}), A^2\mathbb{I}(\mathbb{S}), \ldots, A^{n-1}\mathbb{I}(\mathbb{S})] \\ = & \ \text{Range}[\mathbb{I}(\mathbb{S}), A\mathbb{I}(\mathbb{S})] \end{aligned} \tag{7}$$

where the equality in (7) holds since $A^2 = 0$. Now, (7) implies there exists a vector z such that

$$[\mathbb{I}(\mathbb{S}), A\mathbb{I}(\mathbb{S})]z = \widehat{x}_1. \tag{8}$$

If we break up the set $\mathbb{S}$ into two sets, (i) the set of indexes corresponding to A's first $n - l$ columns, which we denote henceforth by $\mathbb{S}_{1:n-l}$, and (ii) the set of indexes corresponding to A's last l columns, which we denote henceforth by $\mathbb{S}_{n-l+1:n}$, such that $\mathbb{S} = \mathbb{S}_{1:n-l} \cup \mathbb{S}_{n-l+1:n}$, and recall A's definition, we can write the term $A\mathbb{I}(\mathbb{S})$ in (8) as follows:

$$\begin{aligned} A\mathbb{I}(\mathbb{S}) &= \begin{pmatrix} 0 & U' \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbb{I}(\mathbb{S}_{1:n-l}) & 0 \\ 0 & \mathbb{I}(\mathbb{S}_{n-l+1:n}) \end{pmatrix} \\ &= \begin{pmatrix} 0 & U'\mathbb{I}(\mathbb{S}_{n-l+1:n}) \\ 0 & 0 \end{pmatrix} \end{aligned} \tag{9}$$

where U' is, per the definition of A, the matrix that is created by stacking d copies of U the one on top of the other. Therefore, using this

definition of U', the term $U'\mathbb{I}(\mathbb{S}_{n-l+1:n})$ in (9) is rewritten as follows:

$$U'\mathbb{I}(\mathbb{S}_{n-l+1:n}) = \begin{pmatrix} U\mathbb{I}(\mathbb{S}_{n-l+1:n}) \\ U\mathbb{I}(\mathbb{S}_{n-l+1:n}) \\ \vdots \\ U\mathbb{I}(\mathbb{S}_{n-l+1:n}) \end{pmatrix} \tag{10}$$

where the term $U\mathbb{I}(\mathbb{S}_{n-l+1:n})$ is repeated d times. Let now z_1 be the vector that is constructed by z by keeping all the elements of z that in (8) multiply the matrix $\mathbb{I}(\mathbb{S})$, and let z_2 be the vector that is constructed by z by keeping all the elements of z that in (8) multiply the nonzero part of the matrix $A\mathbb{I}(\mathbb{S})$, which is stated in (10). Then, due to (9) and (10), the (8) gives

$$\mathbb{I}(\mathbb{S})z_1 + \begin{pmatrix} U\mathbb{I}(\mathbb{S}_{n-l+1:n})z_2 \\ U\mathbb{I}(\mathbb{S}_{n-l+1:n})z_2 \\ \vdots \\ U\mathbb{I}(\mathbb{S}_{n-l+1:n})z_2 \\ 0 \end{pmatrix} = \widehat{x}_1. \tag{11}$$

Moreover, $\widehat{x}_{\mathbb{E}}$, due to its definition, is in the span of the vectors obtained by taking the rows $\kappa m+1, \ldots, \kappa m+m$ of the columns of the reachability matrix $[\mathbb{I}(\mathbb{S}), A\mathbb{I}(\mathbb{S})]$; in particular, since it is $\mathbb{S}\cap\mathbb{E}=\emptyset$, from (11) we get

$$U\mathbb{I}(\mathbb{S}_{n-l+1:n})z_2 = \widehat{x}_{\mathbb{E}} \tag{12}$$

and indeed we have shown that the vector $\widehat{x}_{\mathbb{E}}$ is in the span of at most $|\mathbb{S}|$ columns of U (12). The contradiction is now obtained because assumption A1 tells us that $|\mathbb{S}| < p_1(l)q_1(l)$ while the statement S2 (which we have assumed initially to hold) tells us the opposite. As a result, the truth of statement S2 implies the truth of statement S2′.

In sum, we proved that the statement S1 implies the statement S1′, as well as, that the statement S2 implies the statement S2′ and, as a result, we showed how Problem 1 can be reduced to the (inapproximable in quasi-polynomial time) Problem 2. Moreover, the reduction is made in polynomial time, since all involved matrices are of polynomial size in l.

We complete Theorem 1's proof with the steps below:

1) Recall that Theorem 2 shows that, unless NP∈BPTIME $(n^{\mathrm{poly}\log n})$, no quasi-polynomial time algorithm can distinguish between the statements S1 and S2; this implies that, under the same assumption, no quasi-polynomial time algorithm can distinguish between the statement S1′ and the statement S2′.
2) Since for any $\delta\in(0,1)$ we can take $q_1(l)=2^{\Omega(\log^{1-\delta} l)}$ in Theorem 2, this implies that the smallest number of inputs rendering $[\mathbf{1}^\top_{m\,d}, \mathbf{0}^\top_{n-dm}]$ reachable cannot be approximated within a multiplicative factor of $q_1(l)$. Indeed, any algorithm which gives an approximation of the smallest number of inputs with a multiplicative factor smaller than $q_1(l)$ would make it possible to distinguish between case S1′ and case S2′. By Theorem 2, the inapproximability factor $q_1(l)$ grows as $2^{\Omega(\log^{1-\delta} l)}$, and since l can be upper and lower bounded by a polynomial in n (since $n\geq l$, and n is at most polynomial in l), we set $\Delta_2(n)=2^{\Omega(\log^{1-\delta} n)}$ in the statement of Theorem 1.
3) Since $\Delta(l)$ is a polynomial in l, as well as, $l\leq n$, we may replace $\Delta(l)$ by some polynomial $\Delta_1(n)$, as in the statement of Theorem 1.

VI. Concluding Remarks

We focused on the minimal reachability Problem 1, which is a fundamental question in optimization and control with applications such as power systems and neural circuits. By exploiting the connection to the variable selection Problem 2, we proved that Problem 1 is hard to approximate. Future work will focus on properties for the system matrix A so that Problem 1 is approximable in polynomial time.

We conclude with an open problem. As we have discussed, the minimum reachability problem is $(0, 2^{\log n})$-approximable by the algorithm which actuates every variable; but $(0, 2^{O(\log^{1-\delta} n)})$ is impossible for any positive δ. We wonder, therefore, whether the minimum number of actuators can be approximated to within a multiplicative factor of say, $\sqrt{n}$ in polynomial time, or, more generally, n^c for some $c\in(0,1)$. Indeed, observe that since $\sqrt{n}=2^{(1/2)\log n}$, the function $\sqrt{n}$ does not belong to $2^{O(\log^{1-\delta} n)}$ for any $\delta>0$. Thus, the present note does not rule out the possibility of approximating the minimum reachability problem up to a factor of $\sqrt{n}$, or more broadly, n^c for $c\in(0,1)$. We remark that such an approximation guarantee would have considerable repercussions in the context of effective control, as at the moment the best polynomial-time protocol for actuation to meet a reachability goal (in terms of worst-case approximation guarantee) is to actuate every variable.

References

[1] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–73, 2011.
[2] F. Muller and A. Schuppert, "Few inputs can reprogram biological networks," *Nature*, vol. 478, no. 7369, 2011, Art. no. E4.
[3] T. Zhou, "Minimal inputs/outputs for a networked system," *IEEE Control Syst. Lett.*, vol. 1, no. 2, pp. 298–303, Oct. 2017.
[4] A. Clark, B. Alomair, L. Bushnell, and R. Poovendran, "Minimizing convergence error in multi-agent systems via leader selection: A supermodular optimization approach," *IEEE Trans. Automat. Control*, vol. 59, no. 6, pp. 1480–1494, Jun. 2014.
[5] A. Clark, L. Bushnell, and R. Poovendran, "A supermodular optimization framework for leader selection under link noise in multi-agent systems," *IEEE Trans. Automat. Control*, vol. 59, no. 2, pp. 283–296, Feb. 2014.
[6] A. Olshevsky, "Minimal controllability problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 249–258, Sep. 2014.
[7] S. Pequito, S. Kar, and A. P. Aguiar, "A framework for structural input/output and control configuration selection in large-scale systems," *IEEE Trans. Automat. Control*, vol. 61, no. 2, pp. 303–318, Feb. 2016.
[8] 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.
[9] 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.
[10] 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.
[11] Y. Zhao, F. Pasqualetti, and J. Cortés, "Scheduling of control nodes for improved network controllability," in *Proc. IEEE 55th Conf. Decision Control*, 2016, pp. 1859–1864.
[12] S. Pequito, G. Ramos, S. Kar, A. Aguiar, and J. Ramos, "Robust minimal controllability problem," *Automatica*, vol. 82, pp. 261–268, 2017.
[13] V. Tzoumas, K. Gatsis, A. Jadbabaie, and G. J. Pappas, "Resilient monotone submodular function maximization," in *Proc. IEEE 56th Annual Conf. Decision Control*, 2017, pp. 1362–1367.
[14] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Sensor placement for optimal kalman filtering," in *Proc. Am. Control Conf.*, 2016, pp. 191–196.
[15] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Near-optimal sensor scheduling for batch state estimation," in *Proc. IEEE 55th Conf. Decision Control*, 2016, pp. 2695–2702.
[16] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202–210, 2017.
[17] L. Carlone and S. Karaman, "Attention and anticipation in fast visual-inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 3886–3893.
[18] M. Amin and J. Stringer, "The electric power grid: Today and tomorrow," *MRS bulletin*, vol. 33, no. 04, pp. 399–407, 2008.
[19] Z. Liu, A. Clark, P. Lee, L. Bushnell, D. Kirschen, and R. Poovendran, "MinGen: Minimal generator set selection for small signal stability in power systems: A submodular framework," in *Proc. IEEE 55th Conf. Decision Control*, 2016, pp. 4122–4129.

[20] *California Partners for Advanced Transit and Highways*, 2006. [Online]. Available: http://www.path.berkeley.edu/

[21] S. Gu *et al.*, "Controllability of structural brain networks," *Nature Commun.*, vol. 6, 2015, Art. no. 8414.

[22] C. Tu, R. P. Rocha, M. Corbetta, S. Zampieri, M. Zorzi, and S. Suweis, "Warnings and Caveats in Brain Controllability," NeuroImage, vol. 176, pp. 83–91, 2018.

[23] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Minimal reachability problems," in *Proc. IEEE 54th Annu. Conf. Decision Control*, 2015, pp. 4220–4225.

[24] Z. Liu, A. Clark, P. Lee, L. Bushnell, D. Kirschen, and R. Poovendran, "Towards scalable voltage control in smart grid: A submodular optimization approach," in *Proc. 7th Int. Conf. Cyber-Phys. Syst.*, 2016, Art. no. 20.

[25] M. Sviridenko, J. Vondrák, and J. Ward, "Optimal approximation for submodular and supermodular optimization with bounded curvature," in *Proc. 26th Annual ACM-SIAM Symp. Discrete Algorithms*, 2014, pp. 1134–1148.

[26] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[27] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functions – I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.

[28] C.-T. Chen, *Linear System Theory and Design*. 3rd ed., New York, NY, USA: Oxford Univ. Press, Inc., 1998.

[29] D. Foster, H. Karloff, and J. Thaler, "Variable selection is hard," in *Proc. Conf. Learn. Theory*, 2015, pp. 696–709.

On a Relaxation of Time-Varying Actuator Placement

Alex Olshevsky, *Member, IEEE*

***Abstract*—We consider the time-varying actuator placement in continuous time, where the goal is to maximize the trace of the controllability Grammian. A natural relaxation of the problem is to allow the binary {0, 1} variable indicating whether an actuator is used at a given time to take on values in the closed interval [0, 1]. We show that all optimal solutions of both the original and the relaxed problems can be given via an explicit formula, and that, as long as the input matrix has no zero columns, the solutions sets of the original and relaxed problem coincide.**

***Index Terms*—Control of networks, network analysis and control, optimization.**

I. Introduction

We consider the time-varying actuator placement problem: informally, given a differential equation with input, we would like to optimize some controllability-related objective while using few nonzero inputs per time step. This is motivated by scenarios where setting an input to something nonzero at a given time carries a fixed cost that can be much larger than the cost of synthesizing the input itself. Our variation of the problem is "time-varying," in the sense that we allow different inputs to be nonzero at different times; this is in contrast to "fixed" actuator placement problems, where one has to select the same set of actuators to be nonzero across all time.

Formally, our goal is to choose a diagonal matrix $V(t)$ whose entries lie in the binary set $\{0, 1\}$ optimizing some controllability-related properties of the resulting differential equation

$$\dot{x}(t) = Ax(t) + BV(t)u(t).$$

where we assume that $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ is a matrix with no zero columns.[1] The multiplication of the input $u(t)$ by the diagonal matrix $V(t)$ can be thought of as choosing to use only certain actuators. Indeed, if $V_{ii}(t) = 0$, then $u_i(t)$ has no effect on $x(t)$, and the i'th entry of the input is ignored at time t.

Typical controllability-related objectives are usually formulated in terms of the controllability Grammian, which we define[2]

$$W_V = \int_0^T e^{At}BV(t)V(t)^TB^Te^{A^Tt}\ dt.$$

The most natural objective is perhaps to minimize $\mathrm{Tr}(W_V^{-1})$, which is proportional to the average energy to move from the origin to a uniformly random point on the unit sphere (see discussion in [17]). However, this function is often challenging to reason about. For example, as a consequence of [16], [23] a number of optimization problems involving $\mathrm{Tr}(W_V^{-1})$ are NP-hard.

We follow several recent papers which instead consider maximization of $\mathrm{Tr}(W_V)$. This is because $\mathrm{Tr}(W_V)$ is easier to reason about and can be used to construct bounds on $\mathrm{Tr}(W_V^{-1})$ (see discussion in [10], [14], [15]). Furthermore, we will seek to do so in the presence of an upper bound on the number of actuators used per unit time step. More formally, denoting $V(t) = \mathrm{diag}(v(t))$, it is typically assumed that the diagonal entries $v_i(t) \in \{0, 1\}$ satisfy the constraint

$$\int_0^T \sum_{i=1}^m |v_i(t)|\ dt \leq \alpha,$$

for some α. We will refer to functions $v_i(t)$ satisfying these constraints as *feasible*. Note that, because we have constrained $v_i(t) \in \{0, 1\}$, this is the same as requiring that

$$\sum_{i=1}^m \mu(\{t : v_i(t) = 1\}) \leq \alpha,$$

where $\mu(\cdot)$ denotes the Lebesgue measure. We will naturally assume that $\alpha \in (0, mT)$, as otherwise the problem is trivial.

A natural relaxation of the problem is to allow each $v_i(t)$ to lie in $[0, 1]$ instead of requiring it to take on the binary values $\{0, 1\}$. We will refer to this as the *relaxed time-varying actuator placement problem*, and the version where $v_i(t)$ are

Manuscript received February 11, 2020; revised March 18, 2020; accepted April 4, 2020. Date of publication April 23, 2020; date of current version May 12, 2020. This research was supported by ARO under Grant W911NF-18-1-0072. Recommended by Senior Editor S. Tarbouriech.

The author is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, and also with the Division of Systems Engineering, Boston University, Boston, MA 02215 USA (e-mail: alexols@bu.edu).

Digital Object Identifier 10.1109/LCSYS.2020.2990099

[1] If B does have zero columns, then the corresponding entry of $u(t)$ does not affect $x(t)$. Consequently, we can simply delete the nonzero columns of B and reindex the vector $u(t)$.

[2] It would be more standard to replace t by $T - t$ in the definition of the controllability Grammian, but since that definition is equivalent to the one we give with a "flipped" $V(t)$, we prefer to avoid dealing with $T - t$'s throughout this letter.

required be in $\{0, 1\}$ will be referred to as the *original time-varying actuator placement problem*. These definitions lead to the main question which is the concern of this letter, namely *understanding when the optimal solutions sets of the original and relaxed problems coincide.*

Compelling motivation for both problems also from the dual formulation, where one considers instead the trace of the observability Gramian. Since the trace of the inverse of the observability Grammian measures the performance of the least-squares estimate of the initial condition over a fixed time-horizon (e.g., [5, p. 21]), our problem formulation can be motivated as "sensor scheduling" where one seeks to achieve the best performance while using as few sensors as possible.

A. Previous Work

This letter is most closely related to the recent work [10], where the same question was considered. We next give a statement of the main results of [10].

Let us adopt the notation b_j for the j'th column of the matrix B. Further, for $i = 1, \ldots, m$, we consider the functions

$$f_i(t) = b_i^T e^{A^T t} e^{At} b_i. \tag{1}$$

It is then possible to give a condition in terms of the functions $f_i(t)$ for the relationship between the original and relaxed time-varying actuator placement problems.

Theorem 1 [10]: The optimal solution set of the relaxed problem is non-empty. Further, assuming that $f_i(t)$ is not constant for all $i \in \{1, \ldots, m\}$, we have that:

1) All optimal solutions of the relaxed problem take $\{0, 1\}$ values almost everywhere.
2) The optimal solution of the solution sets of the original and relaxed actuator scheduling problems coincide.

This theorem is an amalgamation of [10, Ths. 1–3]. It provides an answer to the motivating concern of the present paper. Related theorems in more general settings were also proved in [9] and [11]. However, the condition that $f_i(t)$ are not constant is only shown to be sufficient in this theorem.

It is natural to observe that in discrete-time the optimal schedule for the original problem can be found by a greedy method (see [7, Th. 5]). This suggests it may be possible to give a characterization of the optimal schedule for the original & relaxed problems in the continuous-time model studied here.

This letter is also related to the works [14], [15] which studied combinatorial implications of maximization of the trace of the controllability Grammian, relating them to quantities like centrality and communicability in graphs. Also related is [3] which studied combinatorial aspects of the smallest eigenvalue of the controllability Grammian, which is a measure of the maximum control energy to go from the origin to a point on the unit sphere. Moreover, [2] studied a convex relaxation of the problem of maximizing the smallest eigenvalue of the controllability Gramian. Finally, [9] studied the optimization aspects of finding good actuator schedules with a reformulation of the cost function.

Beyond that, the time-varying actuator placement problem is quite old; for example, a version of it dates back to a paper of Athans in 1972 [1]. There is quite a bit of recent work on understanding efficient algorithms as well as fundamental limitations for this problem. For example, fundamental limitations in terms of unavoidably large control energy have been studied in [17], [19], [26] among others. Algorithms for actuator placement, in either the fixed or time-varying regime, based on randomized sampling [4], [8], [12], [20], convex relaxation [22], [24], or greedy methods [6], [21], [25], [27] were studied in recent works. Given the relatively large amount of work done on different versions of the problem which are not directly related to our motivating concern, we refer the reader to the above papers for a broader overview of the field.

B. Our Contribution

We show that, under our assumption that B has no zero columns, the optimal solution sets of the original and relaxed problems always coincide. This comes out as a byproduct of an explicit formula for the solution of the relaxed problem. In turn, this is done by drawing a connection to (a modification) of the classical notion of a rearrangement of a function.

II. Statement of the Main Result

A. The (Asymmetric) Rearrangement

We need to introduce several concepts and notations to state our main result. We adopt the standard notation that the indicator function $\mathbb{1}_{\mathcal{X}}(x)$ equals one if the point x belongs to the set $\mathcal{X}$ and zero otherwise. Given a Lebesgue measurable subset $\mathcal{X} \subset \mathbb{R}$ of the real line of finite measure, we define its rearrangement $\mathcal{X}^*$ to be the interval $[0, l]$ whose length l is the same as the Lebesgue measure of $\mathcal{X}$. As already mentioned, we adopt the convention of using $\mu(\mathcal{X})$ to denote the Lebesgue measure of the set $\mathcal{X}$.

Given a measurable nonnegative function $f : [0, a] \to \mathbb{R}$ with bounded range, its rearrangement $f^* : [0, a] \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$f^*(x) = \int_0^\infty \mathbb{1}_{\{y \in [0,a]; f(y) > t\}^*}(x)\, dt \tag{2}$$

Intuitively, the rearrangement is that it corresponds to "sorting" the function $f(x)$. In particular [13]:

Proposition 1: The rearrangement $f^*(x)$ is measurable, non-increasing and its level sets have the same measure as the level sets of $f(x)$, i.e., for all $\alpha \in \mathbb{R}$,

$$\mu(\{f(x) \geq \alpha\}) = \mu(\{f^*(x) \geq \alpha\})$$
$$\mu(\{f(x) = \alpha\}) = \mu(\{f^*(x) = \alpha\}).$$

This proposition provides intuition for the notion of a rearrangement: because $f^*(x)$ has the same level sets as $f(x)$ while being non-increasing, it can be thought of as a "sorted" version of $f(x)$. We illustrate this with an example.

Example 1: Consider $f(x) = x^2$ defined on the domain $[0, 1]$. In that case, $\{y \in [0, 1] : f(y) > t\}$ is the set $\{y \in [0, 1] : y^2 > t\}$ which, for $t \in [0, 1]$, equals $(\sqrt{t}, 1]$; when $t > 1$, the set $\{y \in [0, 1] : f(y) > t\}$ is empty. The rearrangement of $(\sqrt{t}, 1]$ is $[0, 1 - \sqrt{t}]$. Thus, for any $x \in [0, 1]$ we have that,

$$f^*(x) = \int_0^1 \mathbb{1}_{[0,1-\sqrt{t}]}(x)\, dt = \int_0^{(1-x)^2} 1\, dt = (1 - x)^2.$$

It is, of course, immediate that $f(x) = x^2$ and $f^*(x) = (1-x)^2$, when defined over the domain $[0, 1]$, have level sets of the same measure.

B. Statement of the Main Result

Our first step is to take the functions $f_i : [0, T] \to \mathbb{R}, i = 1, \ldots, m$, defined in Eq. (1) and define their "concatenation" $F(t) : [0, mT] \to \mathbb{R}$ which consists on putting these functions "side by side" on the interval $[0, mT]$. Formally,

$$F(t) = f_i(t) \text{ when } t \in [(i-1)T, iT).$$

Since the functions $f_i(t)$ are clearly continuous, they have bounded range over $[0, T]$. Further, these functions are clearly nonnegative. As a result, $F(t)$ is also nonnegative and also has bounded range, and therefore the rearrangement $F^*(t)$ is well-defined.

Definition 1: We will say that a function $g : [0, a] \to \mathbb{R} \cup \{+\infty\}$ is strictly decreasing to the right at $x_0 \in (0, a)$ if $g(x_0) > g(x_0 + \epsilon)$ for all ϵ small enough. Similarly, we will say that g is strictly decreasing from the left at $x_0 \in (0, a)$ if $g(x_0 - \epsilon) > g(x_0)$ for all ϵ small enough.

The main result of this letter is the following theorem.

Theorem 2:

1) Suppose $F^*(x)$ is strictly decreasing to the right at $x = \alpha$. Then the unique[3] optimal solution to the relaxed time-varying actuator placement problem is
$$v_i^{\text{opt,r}}(t) = \mathbb{1}_{\{\tau : f_i(\tau) \geq F^*(\alpha)\}}(t).$$
2) Suppose $F^*(x)$ is strictly decreasing from the left at $x = \alpha$. Then the unique optimal solution to the relaxed time-varying actuator placement problem is
$$v_i^{\text{opt, l}}(t) = \mathbb{1}_{\{\tau : f_i(\tau) > F^*(\alpha)\}}(t).$$
3) Suppose $F^*(x)$ is not strictly decreasing at $x = \alpha$ either from the left or to the right. Then the set of optimal solutions to the relaxed time-varying actuator placement problem has more than one element. However, all optimal solutions of the relaxed problem can be parametrized as
$$v_i^{\text{opt, nlr}}(t) = \mathbb{1}_{\{\tau : f_i(\tau) > F^*(\alpha)\}}(t) + \mathbb{1}_{S_i}(t),$$
for some sets
$$S_i \subset \{\tau : f_i(\tau) = F^*(\alpha)\},$$
with
$$\sum_{i=1}^{m} \mu(S_i) = \alpha - \mu(\{\tau : F(\tau) > F^*(\alpha)\}).$$

Since all the solutions exhibited in this theorem are binary, the immediate implication is that the solutions of the original and relaxed problem are always the same. In particular, the condition that the functions $f_i(t)$ not be constant in Theorem 1 is not necessary, once the pathological cases where B has a zero column are ruled out (to see the two conditions are not equivalent, consider $A = 0, B = I$).

This theorem shows that we can write down the optimal solution(s) of the relaxed problem via an explicit formula. Moreover, once the notion of the rearrangement has been introduced, the theorem is quite intuitive: informally, it says that we have to take the "top slice" of the functions $f_1(t), \ldots, f_m(t)$ after sorting. This makes sense if one things of $f_i(t)$ as an "instantaneous contribution" that comes from actuating variable i at time $T - t$.

[3]Of course, we can modify any solution on a set of measure zero without any effect. Thus, here and throughout the remainder of this letter, whenever we refer to equality of functions, we mean up to sets of zero measure.

III. Proof of the Main Result

In this section, we will prove our main result, Theorem 2. We begin by stating several properties of the rearrangement which we will find useful as propositions.

The propositions below hold for all functions $f(x)$ and $g(x)$ such that their rearrangements can be defined, i.e., these functions must be nonnegative and have bounded range. Their proofs are fairly standard. Indeed, it is common in the literature to deal with the "symmetric non-increasing rearrangement" in which $f^*(x)$ is further constructed to be symmetric. For such a notion of rearrangement, the proofs of these facts appear in many places; a standard reference is [13, Ch. 3], which provides hints for many of these, and which can be consulted for general background. We thus do not provide the proofs of these propositions; for completeness, they may be found in the version of this letter on the arxiv [18].

Proposition 2 (Conservation of the L^1 Norm):

$$\int_0^a f(x)\, dx = \int_0^a f^*(x)\, dx.$$

Proposition 3 (Hardy-Littlewood Inequality):

$$\int_0^a f(x)g(x)\, dx \leq \int_0^a f^*(x)g^*(x)\, dx$$

Moreover, we have equality if and only if for almost all s, t, we have that

$$\mu(\{x : f(x) \geq t\} \cap \{x : g(x) \geq s\}) = \min(\mu(\{x : f(x) \geq t\}, \mu(\{x : g(x) \geq s\})).$$

Proposition 4 (Monotonicity): If $f(x) \leq g(x)$ for all x, then $f^*(x) \leq g^*(x)$ for all x. In particular, since a constant function is the rearrangement of itself, if $f(x) \leq 1$ for all $x \in [0, a]$, then $f^*(x) \leq 1$ for all $x \in [0, a]$.

Proposition 5 (Integral Identity): Suppose $b < a$. Then, if $f^*(x)$ is strictly decreasing to the right at $x = b$, we have that

$$\int_0^a f(x)\mathbb{1}_{\{\tau : f(\tau) \geq f^*(b)\}}(x)\, dx = \int_0^b f^*(x)\, dx$$

On the other hand, if $f^*(x)$ is strictly decreasing from the left at $x = b$, then

$$\int_0^a f(x)\mathbb{1}_{\{\tau : f(\tau) > f^*(b)\}}(x)\, dx = \int_0^b f^*(x)\, dx.$$

Proposition 6 (Integral Identity): Let (b_l, b_u) be the largest open interval containing b on which the function $f^*(x)$ is constant. Then if S is a subset of the set $\{x : f(x) = f^*(b)\}$,

then

$$\int_0^a f(x)\left(\mathbb{1}_{\{\tau:f(\tau)>f^*(b)\}}(x)+\mathbb{1}_S(x)\right)\,dx = f^*(b)\mu(S)+\int_0^{b_l} f^*(x).$$

We now turn to our first lemma, which introduces some notation pertaining to functions $f^*(x)$ which are not strictly decreasing from either direction at a point.

Lemma 1: Suppose $f:[0,a]\to\mathbb{R}$ and $f^*(x)$ is not decreasing from either the left or the right at the point $x=b$ with $b\in(0,a)$. Then there exists an open containing b, which we denote by (b^l,b^r), such that

1) $f^*(x)$ is constant on this interval.
2) (b^l,b^r) is the largest open interval containing b with this property.
3)

$$\mu(\{x:f(x)>f^*(b)\}) = b^l$$
$$\mu(\{x:f(x)=f^*(b)\}) = b^r-b^l$$
$$\mu(\{x:f(x)\geq f^*(b)\}) = b^r$$

Proof: Since $f^*(x)$ is nonincreasing, it is immediate that if it is not strictly decreasing from the left or to the right at $x=b$, then it must be constant on some interval (l,r) containing b. We can then define a^l to be infimum of all l such that (l,r) is an interval containing b on which $f^*(x)$ is constant; defining b^r similarly, we obtain that (b^l,b^r) is the largest open interval containing b where $f^*(x)$ is constant. This proves parts (1) and (2).

For part (3), we have that by Proposition 3,

$$\begin{aligned}\mu(\{x:f(x)>f^*(b)\}) &= \mu(\{x:f^*(x)>f^*(b)\})\\ &= \mu([0,b^l)) \text{ or } \mu([0,b^l])\\ &= b^l\end{aligned}$$

and the proof of the second and third identities of item (3) proceed similarly. ∎

Our next lemma is a straightforward generalization, to the continuous space, of the fact that the largest convex combination of a set of numbers (subject to constraint on how big the weights can be) puts as much weight as possible on the largest numbers. We present it without proof.

Lemma 2: Let $g(t):[0,a]\to\mathbb{R}$ be a nonincreasing function and suppose $b\in(0,a)$. Then,

$$\max_{\gamma(t)\in[0,1],\int_0^a\gamma(t)\ dt=b}\ \int_0^a \gamma(t)g(t)\ dt = \int_0^b g(t)\ dt.$$

Moreover,

1) If $g(x)$ is either decreasing to the right or from the left at $x=b$, then the maximum is uniquely achieved by the function $\gamma(t)=\mathbb{1}_{[0,b]}(t)$.
2) If $g(x)$ is not decreasing from the right or the left at $t=b$, let (b^l,b^r) be the open interval guaranteed by Lemma 1. Then the functions which achieve the maximum are

$$\gamma^{\text{opt}}(t) = \mathbb{1}_{[0,b^l)\}}(t)+\lambda(t)\mathbb{1}_Q(t),$$

where $Q\subset[b^l,b^r]$, and $\int_Q\lambda(t)\ dt = b-b^l$.

We next exploit Lemma 2 by applying it to the rearrangement of a function $f^*(x)$, which of course is nonincreasing. The result is stated as the following lemma.

Lemma 3: Let a,b be scalars satisfying $b\leq a$ and let us define $\mathcal{B}$ as the set of functions $\beta(t):[0,a]\to[0,1]$ with

$$\int_0^a \beta(t)\ dt\leq b.$$

Then

$$\max_{\beta\in\mathcal{B}}\int_0^a\beta(t)f(t)\ dt = \int_0^b f^*(t)\ dt. \tag{3}$$

Moreover:

1) If $f^*(t)$ is strictly decreasing to the right at $t=b$, then the unique $\beta(t)$ which achieves this maximum is $\beta(t)=\mathbb{1}_{\{\tau:f(\tau)\geq f^*(b)\}}$.
2) If $f^*(t)$ is strictly decreasing from the left at $t=b$, then the unique $\beta(t)$ which achieves this maximum is $\beta(t)=\mathbb{1}_{\{\tau:f(\tau)>f^*(b)\}}(t)$.
3) If $f^*(t)$ is neither strictly decreasing from the left nor to the right at $t=b$, then the $\beta(t)$ which take values in $\{0,1\}$ almost everywhere which achieve this maximum are

$$\beta(t)=\mathbb{1}_{\{\tau:f(\tau)>f^*(b)\}}(t)+\mathbb{1}_Q(t),$$

where $Q\subset\{\tau:f(\tau)=f^*(b)\}$ with $\mu(Q)=b-\mu(\{\tau:f(\tau)>f^*(b)\})$.

Proof: We prove the lemma in the case $f^*(t)$ is strictly decreasing to the right at $t=b$; the other cases are similar. Observe we only need to prove Eq. (3) with an inequality rather than equality, since by Proposition 5, the function $\beta(t)=\mathbb{1}_{\{\tau:f(\tau)\geq f^*(b)\}}$ makes the left-hand side of Eq. (3) equal to the right-hand side.

Using Proposition 3 we have that for any $\beta(t)\in\mathcal{B}$,

$$\int_0^a\beta(t)f(t)\ dt\leq\int_0^a\beta^*(t)f^*(t)\ dt. \tag{4}$$

By Proposition 3 we have that β^* integrates to b over $[0,a]$ just like β. By Proposition 4, we have that $\beta^*(t)\leq 1$. Moreover, since the rearrangement of a function is always nonnegative as an immediate consequence of Eq. (2), we also have that $\beta^*(t)$ is nonnegative. Thus:

$$\int_0^a\beta(t)f(t)\ dt\leq\max_{\gamma(t)\in[0,1],\ \int_0^a\gamma(t)\leq b}\ \int_0^a\gamma(t)f^*(t)\ dt.$$

However, since $f^*(t)$ is nonincreasing by Proposition 4, by Lemma 2 we must have that $\gamma^{\text{opt}}(t)=1_{[0,b]}(t)$. Thus

$$\int_0^a\beta(t)f(t)\ dt\leq\int_0^a\mathbb{1}_{[0,b]}(t)f^*(t)\ dt.$$

Since this holds for all $\beta\in\mathcal{B}$, we have proved Eq. (3).

It remains to characterize what $\beta(t)$ achieve equality in Eq. (3). For this, all the inequalities in the proof we've just given must be satisfied with equality. We next go through several of these inequalities and discuss how having equality in them constrains $\beta(t)$.

First, by Lemma 2 the optimal $\gamma(t)$ is unique and equals $\mathbb{1}_{[0,b]}(t)$, so we must have $\beta^*(t)=\mathbb{1}_{[0,b]}(t)$. In particular, this implies that $\beta(t)$ is the indicator function of a set of Lebesgue measure b. Let S denote that set.

Second, applying the equality conditions of Proposition 3 to Eq. (4), the set S is such that for almost all s, t,

$$\mu(\{x : f(x) \geq t\} \cap \{x : \mathbb{1}_S(x) \geq s\})$$

and

$$\min(\mu(\{x : f(x) \geq t\}, \mu(\{x : \mathbb{1}_S(x) \geq s\})$$

are the same. But this implies that

$$\mu(\{x : f(x) \geq t) \cap S) = \min(\mu(\{x : f(x) \geq t\}), \mu(S))$$

for almost all t. The last statement implies that, for almost all t, the level set $\{x : f(x) \geq t\}$ either contains or is contained in S (up to a set of zero measure).

It is tempting to argue that since $\mu(S) = b$ and $\mu\{x : f(x) \geq f^*(b)\} = b$ (because $f^*(x)$ is decreasing to the right at $x = b$), and one of these two sets contains the other, they must be equal (again up to a set of measure zero). Unfortunately, it might be that the choice $t = f^*(b)$ is not included in the "almost all" t above. Thus we proceed as follows. We consider the level sets $L_\epsilon = \{x : f(x) \geq f^*(b) - \epsilon\}$ with $\epsilon > 0$ which satisfy $\mu(L_\epsilon) \geq b$. Since the measure of S is exactly b, it follows that $S \subset L_\epsilon$, up to a set of measure zero, for almost all ϵ. Since the sets L_ϵ are nested, S is in fact contained in every L_ϵ for $\epsilon > 0$. So S is contained in the intersection of $\{L_\epsilon, \epsilon > 0\}$, which is $\{x : f(x) \geq f^*(b)\}$. Since, as mentioned above, the last set has the same measure as S, we conclude it equals S up to a set of measure zero. ∎

With this last lemma in place, we can now turn to the proof of our main result.

Proof of Theorem 2: We first argue that the solutions we have proposed are feasible. First, suppose that $F^*(x)$ is strictly decreasing to the right at $x = \alpha$. Then, using Proposition 3, we have

$$\begin{aligned}\sum_{i=1}^m \int_0^T |v_i^{\text{opt, r}}(t)|\ dt &= \sum_{i=1}^m \mu(\{\tau : f_i(\tau) \geq F^*(\alpha)\}) \\ &= \mu(\{\tau : F(\tau) \geq F^*(\alpha)\}) \\ &= \mu(\{\tau : F^*(\tau) \geq F^*(\alpha)\}) \\ &= \alpha,\end{aligned}$$

where the last step follows from the assumption that F^* is decreasing to the right at α. The case when $F^*(x)$ is strictly decreasing from the left at $x = \alpha$ follows by an almost identical argument.

Next, suppose $F^*(x)$ is not strictly decreasing either from the left or to the right at $x = \alpha$. Then

$$\begin{aligned}\sum_{i=1}^m \int_0^T |v_i^{\text{opt, nlr}}(t)|\ dt &= \sum_{i=1}^m \mu(\{\tau : f_i(\tau) > F^*(\alpha)\}) + \mu(S) \\ &= \mu(\{\tau : F^*(\tau) > F^*(\alpha)\}) \\ &\quad + (\alpha - \mu(\{\tau : F^*(\tau) > F^*(\alpha)\})) \\ &= \alpha,\end{aligned}$$

Since it is immediate that all $v_i^{\text{opt, r}}(t), v_i^{\text{opt, l}}(t), v_i^{\text{opt, nlr}} \in [0, 1]$, we conclude that in all cases the proposed solutions are feasible.

Let us adopt the notation $J_{v^{\text{opt, r}}}$ for the cost corresponding to the functions $v_i^{\text{opt, r}}$. Let us compute this cost under the assumption that $F^*(t)$ is decreasing from the right at $t = \alpha$. We have that

$$\begin{aligned}J_{v^{\text{opt, r}}} &= \text{Tr} \int_0^T \sum_{i=1}^m e^{At} b_i \mathbb{1}_{\{\tau : f_i(\tau) \geq F^*(\alpha)\}}(t) b_i^T e^{A^T t}\ dt \\ &= \int_0^T \sum_{i=1}^m \mathbb{1}_{\{\tau : f_i(\tau) \geq F^*(\alpha)\}}(t) b_i^T e^{A^T t} e^{At} b_i\ dt \\ &= \int_0^T \sum_{i=1}^m \mathbb{1}_{\{\tau : f_i(\tau) \geq F^*(\alpha)\}}(t) f_i(t)\ dt \\ &= \int_0^{mT} F(t) \mathbb{1}_{\{\tau : F(\tau) \geq F^*(\alpha)\}}(t)\ dt \\ &= \int_0^\alpha F^*(t)\ dt, \end{aligned} \tag{5}$$

and the last step used Proposition 5 and the assumption that $F^*(t)$ is strictly decreasing to the right at $t = \alpha$. Thus we have shown that the choice of functions $v_i^{\text{opt, r}}(t)$ achieves a cost of $\int_0^\alpha F^*(t)\ dt$. The case when $F^*(t)$ is strictly decreasing from the left at $t = \alpha$ follows by an almost identical argument.

We next argue that, under the assumption that $F^*(t)$ is decreasing neither from the left nor the right, the functions $v_i^{\text{opt, nlr}}$ achieve the same cost. Indeed, let (α^l, α^r) be the largest open interval containing α on which $F^*(t)$ is non-decreasing whose existence was guaranteed by Lemma 1. Note that lemma tells us that $\mu\{\tau : F^*(\tau) > F^*(\alpha)\} = \alpha^l$.

Define $S_i' = S_i + (i-1)T$, where S_i comes from the theorem statament and we translate it by $(i-1)T$ to make it so that $S_i' \subset [(i-1)T, iT]$. Further defining $S = \cup_{i=1}^m S_i'$ and proceeding similarly as before,

$$\begin{aligned}J_{v^{\text{opt, nlr}}} &= \int_0^{mT} F(t)\big(\mathbb{1}_{\{\tau : F(\tau) > F^*(\alpha)\}}(t) + \mathbb{1}_S(t)\big)\ dt \\ &= \int_0^{\alpha^l} F^*(t)\ dt + \mu(S) F^*(\alpha) \\ &= \int_0^{\alpha^l} F^*(t)\ dt + (\alpha - \mu\{F(\tau) > F^*(\alpha)\}) F^*(\alpha) \\ &= \int_0^{\alpha^l} F^*(t)\ dt + (\alpha - \alpha^l) F^*(\alpha) \\ &= \int_0^\alpha F^*(t)\ dt,\end{aligned}$$

where we relied on Proposition 3 in the second step. To summarize, we have shown that under the appropriate assumptions, all three of $v_i^{\text{opt, r}}(t)$ and $v_i^{\text{opt, l}}(t)$ and $v_i^{\text{opt, nlr}}(t)$ achieve the cost of $\int_0^\alpha F^*(t)\ dt$.

We next show that, for any choice of functions $v_i(t)$, we will attain a cost that is upper bounded by $\int_0^\alpha F^*(t)\ dt$.

Let us adopt the notation J_v for the cost corresponding to the functions $v_i(t), i = 1, \ldots, m$; our goal is to show that, as long as $v_i(t)$ are feasible, we have $J_v \leq \int_0^\alpha F^*(t)\ dt$.

Indeed, for any feasible functions $v_i(t)$ we have that $v_i(t) \in [0, 1]$ which means that

$$J_v = \mathrm{Tr} \int_0^T \sum_{i=1}^m v_i^2(t) e^{At} b_i b_i^T e^{A^T t} \, dt \tag{6}$$

$$\leq \mathrm{Tr} \int_0^T \sum_{i=1}^m v_i(t) e^{At} b_i b_i^T e^{A^T t} \, dt \tag{7}$$

$$= \sum_{i=1}^m \int_0^T v_i(t) f_i(t) \, dt = \int_0^{mT} q(t) F(t) \, dt \tag{8}$$

where we define

$$q(t) = \sum_{i=1}^m v_i(t - (i-1)T) \mathbb{1}_{[(i-1)T, iT)}(t)$$

Observing that

$$\int_0^{mT} |q(t)| \, dt = \int_0^T \sum_{i=1}^m |v_i(t)| \, dt \leq \alpha,$$

by feasibility of $v_i(t)$, we can now apply Lemma 3 to Eq. (8) and, by Eq. (5), obtain $J_v \leq \int_0^\alpha F^*(t) \, dt$.

Putting it all together, we have thus shown that, under the appropriate assumptions, each of $v_i^{\mathrm{opt},1}(t), v_i^{\mathrm{opt,r}}(t), v_i^{\mathrm{opt,\,nlr}}(t)$ are optimal. It remains to prove that these are the only optimal choices. For this, we must analyze the cases of equality in the above bounds.

Observe that to achieve equality we need to have equality starting from the Eq. (6) through the end of the proof. In particular, we must have equality in the application of Lemma 3. But that lemma spells out conditions for equality. In particular, Lemma 3 forces $q(t) = \mathbb{1}_{\{\tau : F(\tau) \geq F^*(\alpha)\}}(t)$ when $F^*(t)$ is decreasing from the right at α and $q(t) = \mathbb{1}_{\{\tau : F(\tau) > F^*(\alpha)\}}(t)$ when it is decreasing from the left. By definition of $q(t)$, this is the same as having $v_i(t) = \mathbb{1}_{\{\tau : f_i(\tau) \geq F^*(\alpha)\}}(t)$ and $v_i(t) = \mathbb{1}_{\{\tau : f_i(\tau) \geq F^*(\alpha)\}}(t)$. This concludes the proof for the case where $F^*(t)$ is either strictly decreasing to the right or from the left at $t = \alpha$.

It only remains to analyze the cases of equality in the case where $F^*(t)$ is not strictly decreasing either from the left or to the right at $t = \alpha$. First, observe that because B has no zero columns and e^{At} is always nonsingular, we have that $\mathrm{Tr}(e^{At} b_i b_i^T e^{A^T t}) > 0$, for all $i \in \{1, \ldots, m\}$ and $t \in [0, T]$. In particular, because $v_i(t) \in [0, 1]$, the implication of this is that to achieve equality going from Eq. (6) to Eq. (7), we must have that each $v_i(t) \in \{0, 1\}$ almost everywhere. This implies the function $q(t)$ must be binary as well.

Having established that, we apply the conditions for equality in the last item of Lemma 3. That lemma tells us that we must have $q(t) = \mathbb{1}_{\{\tau : F(\tau) > F^*(\alpha)\}}(t) + \mathbb{1}_S$, for a subset S of the set $\{t : F(t) = F^*(\alpha)\}$ with $\mu(S) = \alpha - \mu(\{\tau : F^*(\tau) > F^*(\alpha)\})$. This concludes the proof. ∎

REFERENCES

[1] M. Athans, "On the determination of optimal costly measurement strategies for linear stochastic systems," *Automatica*, vol. 8, no. 4, pp. 397–412, 1972.

[2] G. Baggio, S. Zampieri, and C. W. Scherer, "Gramian optimization with input-power constraints," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, 2019, pp. 5686–5691.

[3] N. Bof, G. Baggio, and S. Zampieri, "On the role of network centrality in the controllability of complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 643–653, Sep. 2017.

[4] S. D. Bopardikar, "Sensor selection via randomized sampling," 2017. [Online]. Available: arXiv:1712.06511.

[5] S. Boyd, "Introduction to linear dynamical systems," Lecture Notes EE263, Stanford University, Stanford, CA, USA, 2008.

[6] P. V. Chanekar, N. Chopra, and S. Azarm, "Optimal actuator placement for linear systems with limited number of actuators," in *Proc. Amer. Control Conf. (ACC)*, 2017, pp. 334–339.

[7] A. S. A. Dilip, "The controllability Gramian, the hadamard product, and the optimal actuator/leader and sensor selection problem," *IEEE Control Syst. Lett.*, vol. 3, no. 4, pp. 883–888, Oct. 2019.

[8] A. Hashemi, M. Ghasemi, H. Vikalo, and U. Topcu, "A randomized greedy algorithm for near-optimal sensor scheduling in large-scale sensor networks," in *Proc. Ann. Amer. Control Conf. (ACC)*, 2018, pp. 1027–1032.

[9] T. Ikeda and K. Kashima, "On sparse optimal control for general linear systems," *IEEE Trans. Autom. Control*, vol. 64, no. 5, pp. 2077–2083, May 2019.

[10] T. Ikeda and K. Kashima, "Sparsity-constrained controllability maximization with application to time-varying control node selection," *IEEE Control Syst. Lett.*, vol. 2, no. 3, pp. 321–326, Jul. 2018.

[11] T. Ikeda and K. Kashima, "Sparse optimal feedback control for continuous-time systems," in *Proc. 18th Eur. Control Conf. (ECC)*, 2019, pp. 3728–3733.

[12] A. Jadbabaie, A. Olshevsky, and M. Siami, "Deterministic and randomized actuator scheduling with guaranteed performance bounds," 2018. [Online]. Available: arXiv:1805.00606.

[13] E. H. Lieb and M. Loss, "Analysis," in *Graduate Studies in Mathematics*, vol. 14. Providence, RI, USA: American Mathematical Society, 2001.

[14] E. Nozari, F. Pasqualetti, and J. Cortés, "Time-invariant versus time-varying actuator scheduling in complex networks," in *Proc. Amer. Control Conf. (ACC)*, 2017, pp. 4995–5000.

[15] E. Nozari, F. Pasqualetti, and J. Cortés, "Heterogeneity of central nodes explains the benefits of time-varying control scheduling in complex dynamical networks," *J. Complex Netw.*, vol. 7, no. 5, pp. 659–701, 2019.

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

[17] A. Olshevsky, "Eigenvalue clustering, control energy, and logarithmic capacity," *Syst. Control Lett.*, vol. 96, pp. 45–50, Oct. 2016.

[18] A. Olshevsky, "On a relaxation of time-varying actuator placement," 2019. [Online]. Available: arXiv:1912.09454.

[19] 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.

[20] M. Siami and A. Jadbabaie, "Deterministic polynomial-time actuator scheduling with guaranteed performance," in *Proc. Eur. Control Conf. (ECC)*, 2018, pp. 113–118.

[21] T. Summers and M. Kamgarpour, "Performance guarantees for greedy maximization of non-submodular controllability metrics," in *Proc. 18th Eur. Control Conf. (ECC)*, 2019, pp. 2796–2801.

[22] T. Summers and I. Shames, "Convex relaxations and gramian rank constraints for sensor and actuator selection in networks," in *Proc. IEEE Int. Symp. Intell. Control*, 2016, pp. 1–6.

[23] 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.

[24] A. Zare, H. Mohammadi, N. K. Dhingra, T. T. Georgiou, and M. R. Jovanovic, "Proximal algorithms for large-scale statistical modeling and sensor/actuator selection," *IEEE Trans. Autom. Control*, early access, Oct. 18, 2019, doi: 10.1109/TAC.2019.2948268.

[25] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, 78, pp. 202–210, Apr. 2017.

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

[27] Y. Zhao, F. Pasqualetti, and J. Cortés, "Scheduling of control nodes for improved network controllability," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, 2016, pp. 1859–1864.

On the Inapproximability of the Discrete Witsenhausen Problem

Alex Olshevsky

***Abstract*—We consider a discrete version of the Witsenhausen problem where all random variables are bounded and take on integer values. Our main goal is to understand the complexity of computing good strategies given the distributions for the initial state and second-stage noise as inputs to the problem. Following Papadimitriou and Tsitsiklis, who showed that computing the optimal solution is NP-complete, we construct a sequence of problem instances with the initial state uniform over a set of size n and the noise uniform over a set of size at most n^2, such that finding a strategy whose cost is a multiplicative $n^{2-\epsilon}$ approximation to the optimal cost is NP-hard for any $\epsilon > 0$.**

***Index Terms*—Decentralized control, computational complexity.**

I. Introduction

WITSENHAUSEN'S seminal counterexample [2] demonstrated that linear strategies are not always in sequential stochastic control. The counterexample consists of a two-agent optimization problem with what has come to be known as a non-classical information pattern, in that it involves two agents acting in sequence, with the second agent having no knowledge of the information seen by the first agent. In the decades since [2], a considerable literature has sprung up analyzing control problems with non-classical information patterns [3]. Nevertheless, a complete analysis of the Witsenhausen's original counterexample is lacking, though considerable progress has been made in understanding the relation between optimal strategies and information patterns [4], [5], [6], [7], [8], [9].

The goal of this letter is to contribute to the literature which attempts to explain why Witsenhausen's problem is difficult. Our starting point is this letter [1], which considered a discrete version of the Witsenhausen counterexample where all the random variables and controls were restricted to be integers. This problem formulation can be obtained by quantizing the Witsenhausen problem and rescaling [10], [11]. Furthermore, the distribution of the initial state of the system and the noise were viewed as inputs; in Witsenhausen's original formulation, both of these were taken to be Gaussian. It was shown in [1] that computation of the optimal strategy for this version of the Witsenhausen problem is NP-complete.

While such NP-hardness results do not have any implications for Witsenhausen's original counterexample, in the generalized scenario where the initial state and noise have arbitrary distributions, they have a fairly powerful message. Indeed, let us consider what would count as a solution of the Witsenhausen problem in this more general scenario. Presumably, one would want a formula for the optimal strategy as a function of the initial state and noise distributions. However, such a formula would be quite useless if it could not be evaluated efficiently. Thus at the very least there should exist an efficient algorithm for the computation of the optimal strategy, and it is exactly this that [1] rules out.

Our goal in this letter is to strengthen the results of [1]. We seek to address the question of whether it is possible to find approximately optimal solutions to the Witsenhausen problem. It might initially seem that there are reasons to be hopeful. Indeed, the reduction in [1] reduces the Witsenhausen problem to a 3D matching problem, and, although 3D matching is NP-hard, a $4/3+\epsilon$ approximation algorithm is available for any $\epsilon > 0$ [12]. Moreover, constant factor approximation results were derived in [13] for a different, but finite dimensional problem formulation, albeit with Gaussian noises.

Unfortunately, our main result rules out the possibility of a favorable approximation with the discrete Witsenhausen problem with arbitrary initial state and noise distribution. We describe a family of examples, where the initial state is uniform over a set of n integers, and the noise is uniform over a set of at most n^2 integers, and it is NP-hard to find a strategy whose cost is upper bounded by $n^{2-\epsilon}$ times the cost of the optimal strategy, for any $\epsilon > 0$.

One might wonder if the multiplicative $n^{2-\epsilon}$ factor is the best one could do, i.e., if the problem might be even more difficult to approximate than that. In that direction, we show that if the initial distribution has support $\mathcal{X}$ and the noise distribution has support $\mathcal{Z}$, then it is always possible to approximate the optimal Witsenhausen strategy to within a multiplicative factor of $|\mathcal{X}|^3|\mathcal{Z}|^4$. Plugging in $|\mathcal{X}| = n$ and $\mathcal{Z} \leq n^2$ for the construction of the previous paragraph, we obtain that a multiplicative n^{11} approximation is possible in that case. This

Manuscript received February 12, 2019; revised April 10, 2019; accepted April 10, 2019. Date of publication April 18, 2019; date of current version April 30, 2019. This work was supported in part by NSF under Award 1740451, and in part by ARO under Award W911NF-18-1-0072. Recommended by Senior Editor J.-F. Zhang.

The author is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, and also with the Division of Systems Engineering, Boston University, Boston, MA 02215 USA (e-mail: alexols@bu.edu).

Digital Object Identifier 10.1109/LCSYS.2019.2911925

shows a limit to how much one could potentially improve the $n^{2-\epsilon}$-inapproximability result we described above.

The remainder of this letter is organized as follows. Section II contains technical background, including a formal definition of the Witsenhausen problem and the discretizations we described above. Section III contains a proof of the $n^{2-\epsilon}$-inapproximability result while Section IV contains a proof of the $|\mathcal{X}|^3|\mathcal{Z}|^4$ approximation result.

II. Background

We begin with an informal description of the Witsenhausen problem. Two agents attempt to stabilize a system by bringing its state close to zero in two time steps. The first agent observes the initial state X_0, which we assume to be a random variable with a known distribution. The first agent applies the control u_1, so that the state becomes $X_0 + u_1$. Now the second agent can only see a noisy version $X_0 + u_1 + Z$ of the state, where Z is some random variable with a known distribution. It applies a control u_2 which is therefore constrained to be a function of $X_0 + u_1 + Z$. The final cost depends only on the size of the control applied by the first agent as well as the final distance to the origin:

$$E\Big[u_1^2 + K(X_0 + u_1 + u_2)^2\Big],$$

where $K > 0$ is some constant. In particular, the control applied by the second agent is "free."

In the classical Witsenhausen counterexample [2], it is assumed that the initial state is $X_0 \sim N(0, \sigma_0^2)$ while the noise is $Z \sim N(0, 1)$, but in this letter we will consider arbitrary distributions for X_0, Z. We will find it convenient to reformulate the problem in a way that makes the inherent constraints explicit as follows. Given independent random variables X_0, Z we are looking for maps $T : \mathbb{R} \to \mathbb{R}$, $\delta : \mathbb{R} \to \mathbb{R}$ which minimize the cost function

$$E\Big[(T(X_0) - X_0)^2 + K(T(X_0) + \delta(T(X_0) + Z))^2\Big]. \tag{1}$$

Moreover, we will denote this quantity as $\Phi(p_{X_0}, p_Z, T, \delta)$ and refer to it as the "Witsenhausen cost." Furthermore, we will refer to $E[(T(X_0) - X_0)^2]$ as the "first-stage" or "transportation cost," while $E[[T(X_0) + \delta(T(X_0) + Z))]^2]$ will be referred to as the "second-stage" cost.

The discrete Witsenhausen problem, defined formally next, is simply the restriction of this problem to random variables and maps which take on integer values. For convenience, in the sequel we use $\mathbb{Z}$ to denote the set of integers.

Definition 1: Let X_0, Z be independent bounded random variables taking on integer values with probability mass functions p_{X_0}, p_Z. The Witsenhausen problem asks for maps $T : \mathbb{Z} \to \mathbb{Z}$ and $\delta : \mathbb{Z} \to \mathbb{Z}$ achieving the minimum in Eq. (1). We will use $\Phi^*(p_{X_0}, p_Z)$ to refer to the optimal cost as a function of the problem parameters.[1]

This problem was essentially introduced in [10]. It is not hard to see that an optimal solution exists: we can restrict our attention to a finite set of maps T, δ, as there is no need to consider maps which move some values in $\mathcal{X}$ too far. It is also standard that, given T, the corresponding δ can be found by solving a least squares problem.

Our goal in this letter is to prove the following theorem, which refines a result of [1] that the discrete Witsenhausen problem is NP-hard. We will adopt the convention of using $\mathcal{X}$ to refer to the support of X and $\mathcal{Z}$ to refer to the support of Z.

Theorem 1: Consider the discrete Witsenhausen problem restricted to problem instances where $|\mathcal{X}| = n$ and $|\mathcal{Z}| \leq n^2$. Unless $P = NP$, for any $\epsilon > 0$ there does not exist a polynomial-time algorithm which returns a number ϕ satisfying

$$\phi \leq n^{2-\epsilon}\Phi^*(p_{X_0}, p_Z)$$

such that $\phi = \Phi(p_{X_0}, p_Z, T, \delta)$ for some choice of maps T, δ.

We also show that the $n^{2-\epsilon}$ factor in the inapproximability result cannot be improved too much.

Theorem 2: There is a polynomial-time algorithm which returns T, δ satisfying

$$\Phi(p_{X_0}, p_Z, T, \delta) \leq |\mathcal{X}|^3|\mathcal{Z}|^4\Phi^*(p_{X_0}, p_Z)$$

Indeed, plugging in $|\mathcal{X}| = n, |\mathcal{Z}| = n^2$ into this last theorem, we obtain that in the setting described by Theorem 1, this provides an n^{11} multiplicative approximation.

III. Proof of Theorem 1

We now turn to a sequence of lemmas whose culmination will be the proof of Theorem 1. Our starting point is a definition which later on will be key to the way we will definite the initial state X_0.

Definition 2: An integer set S is called a Sidon set of order p if all the sums

$$s_1 + s_2 + \cdots + s_p$$

with $s_1, \ldots, s_p \in S$ and $s_1 \leq s_2 \leq \cdots \leq s_p$ are distinct.

For example, $S = \{1, 2, 4\}$ is a Sidon set of order 2 because the pairwise sums of elements from this set are distinct; but $S = \{1, 2, 3, 4\}$ is not a Sidon set of order 2 because $3 + 3 = 4 + 2$.

It is well-known that Sidon sets of arbitrary order exist and can be easily constructed. We will need the following variation of this fact, which is very similar to a lemma from [1].

Lemma 1: There exists a Sidon set S of order 4 with $|S| = k$ satisfying $S \subset \{1, 2, \ldots, 20k^8\}$. Moreover, it is possible to construct S in polynomial time in k.

Proof: We prove this lemma by induction. When $k = 1$, we can simply choose $S = \{1\}$. Now suppose we have a Sidon set $S = \{s_1, s_2, \ldots, s_k\}$, with s_i being distinct positive integers, and $\max_{i=1,\ldots,k} s_i \leq 20k^8$. We construct a Sidon set of size $k+1$ by choosing a positive integer $s_{k+1} \leq 20(k+1)^8$ to add to S.

To ensure this works, we need that

$$s_a + s_b + s_c + s_d \neq s_e + s_f + s_g + s_{k+1}$$

[1] A common convention in the literature is to specify the distribution of $Y = T(X_0) + Z$ conditioned on $T(X_0)$, but since X_0 and Z are independent here, it is easier in this case to simply specify the distribution of Z.

for all possible choices $a \leq b \leq c \leq d, e \leq f \leq g$ from $a, b, c, d, e, f, g \in \{1, \ldots, k+1\}$. In other words, we need to have

$$s_{k+1} \neq s_a + s_b + s_c + s_d - s_e - s_f - s_g.$$

There are at most $(k+1)^7$ choices of a, b, c, d, e, f, g, and each inequality produces at most one value for s_{k+1} to avoid. Indeed, it is possible for an inequality to produce no value for s_{k+1} to avoid, for example if s_{k+1} cancels from both sides. However, if this does not happen, then the corresponding choice of a, b, c, d, e, f, g results in one value for s_{k+1} to avoid.

It follows that this produces a Sidon set of order 4 if we just place s_{k+1} outside this set of at most $(k+1)^7$ values. We will place s_{k+1} in the range $(20k^8, 20(k+1)^8]$, which we can always do by the pigeonhole principle, since

$$20(k+1)^8 - 20k^8 > 20 \cdot 8k^7 > (2k)^7 \geq (k+1)^7.$$

Finally, we can construct S via the above procedure which clearly takes polynomial time. ∎

The bounds of the lemma are rather loose and it is possible to improve them, but they suffice for our purposes.

Our next step is discuss a variation on the notion of the chromatic number which we will need. Although the following discussion seems unrelated to Sidon sets, we will bring the two concepts together in our NP-hardness proof later on.

The chromatic number of a graph is the minimum number of colors needed to color the vertices so that no adjacent vertices share the same color; we will use $\kappa(G)$ to denote the chromatic number of the graph G. We will need to use a certain notion which we call the l_2-chromatic number, which as far as we know has not been previously considered, and we motivate this notion with the following discussion.

We may formulate the search for the chromatic number of the graph $G = (\{1, \ldots, n\}, E)$ as minimizing

$$\Phi(\gamma) := \max_{i=1,\ldots,n} \gamma(i) - \min_{i=1,\ldots,n} \gamma(i),$$

over all functions $\gamma : V \to \mathbb{Z}$ satisfying $\gamma(i) \neq \gamma(j)$ for all $(i, j) \in E$. Indeed, the objective $\Phi(\gamma)$ is precisely the number of colors needed minus one.

The quantity $\Phi(\gamma)$ may be thought of as a measure of dispersion. This motivated the introduction of the l_2-chromatic number, which uses a slightly different measure of dispersion: the variance of the distribution of $\gamma(i)$ about zero.

Definition 3: Given an undirected graph $G = (\{1, \ldots, n\}, E)$, the l_2-chromatic number asks for a function $\gamma : V \to \mathbb{Z}$ satisfying $\gamma(i) \neq \gamma(j)$ for all $(i, j) \in E$ and minimizing

$$\gamma^* := \frac{1}{n} \sum_{i=1}^{n} \gamma^2(i).$$

We remark that the graph G should not have any self-loops, for otherwise the constraint $\gamma(i) \neq \gamma(j)$ for all edges (i, j) in G is impossible to satisfy.

We next use the concept of Sidon sets to give a way to construct a discrete Witsenhausen problem starting from a graph. The ensuring sequence of lemmas will show that computation of the l_2-chromatic number on that graph will then be equivalent to computation of the optimal Witsenhausen strategy.

Definition 4: Given a graph $G = (\{1, \ldots, n\}, E)$ and an integer B, construct an instance of the discrete Witsenhausen as follows:

- Let $\{y_1, \ldots, y_n\}$ be Sidon set of order 4 with n elements, and set $x_i = l y_i$, where $l = 4(\lceil n^{1.5} \rceil + 1)$. Generate X_0 to be uniform over $x_1, \ldots, x_n$.
- Generate Z to be uniform over all the pairs $(x_i - x_j)/2$, $(i, j) \in E$.
- Set K to be any number strictly bigger than n^5.

Observe that the graph G enters the definition of the corresponding discrete Witsenhausen problem solely through the distribution of Z. Observe further that the support of Z always symmetric about the origin since $(i, j) \in E$ whenever $(j, i) \in E$. Note also that the support of the random variable X_0, i.e., the set $\{x_1, \ldots, x_n\}$, is a Sidon set of order 4 with n elements (because it is obtained via scaling each element of a Sidon set by the same factor l). Finally, note that this construction may be performed in polynomial time in n as a consequence of Lemma 1 which tells us that the set $\{y_1, \ldots, y_n\}$ may be constructed in polynomial time; that all remaining operations take polynomial time is obvious.

The equivalence of l_2-chromatic number on the original graph and the cost of the optimal Witsenhausen strategy in this construction is established in the following two lemmas.

Lemma 2: Suppose $0 \leq B \leq n^2$. If the discrete Witsenhausen problem constructed in Definition 4 has a solution with cost at most B, then the l_2-chromatic number of the graph G is at most B.

Proof: Let T, δ be maps which achieve a cost of at most B in the resulting Witsenhausen problem. We will define

$$\gamma(i) = T(x_i) - x_i,$$

and argue that this choice of γ works. The key observation is that, if the discrete Witsenhausen problem constructed in this way has a cost at most B, then it has zero second-stage cost, i.e., we must have with probability one that

$$T(X_0) = \delta(T(X_0) + Z). \tag{2}$$

This follows because of the way K was chosen. Formally, observe that if Eq. (2) fails with positive probability, then, because X_0, Z were constructed to be uniform over $\mathcal{X}$ and $\mathcal{Z}$, it fails with probability at least $(1/|\mathcal{X}|)(1/|\mathcal{Z}|) \geq 1/n^3$. Moreover, when Eq. (2) fails, then because both the left-hand side and the right-hand side of this equation are integer, it follows they differ by at least one. Thus, in that case the expectation of the Witsenhausen cost of Eq. (1) is at least $(1/n^3)K \cdot 1 > B$. This is a contradiction. We have thus shown that Eq. (2) holds with probability one.

In particular, this means that for all possible $x_i, x_j \in \mathcal{X}$ such that $T(x_i) \neq T(x_j)$, and all possible $z_a, z_b \in \mathcal{Z}$, we must have

$$T(x_i) + z_a \neq T\big(x_j\big) + z_b. \tag{3}$$

Indeed, if Eq. (3) fails, then it is immediate that a zero second-stage cost cannot be obtained.

We now claim that, due to the way X_0 was defined in Definition 4, we can conclude that actually $T(x_i) \neq T(x_j)$ for all pairs $i, j = 1, \ldots, n$, so that the conclusion of the previous paragraph is actually applicable to all pairs i, j. Indeed, suppose $T(x_i) = T(x_j)$ for some pair i, j. Since $|x_i - x_j| > 4n^{1.5}$, we have that either $|T(x_i) - x_i| > 2n^{1.5}$ or $|T(x_j) - x_j| > 2n^{1.5}$. Either one of these will imply the first-stage transportation cost is strictly bigger than n^2 and thus strictly bigger than B.

Putting the last two paragraphs together, we have that for all realizations $x_i, x_j, \in \mathcal{X}, z_i, z_j \in \mathcal{Z}$, we have that

$$T(x_i) + z_a \neq T\big(x_j\big) + z_b.$$

In particular,

$$T(x_i) - T\big(x_j\big) \neq z_b - z_a,$$

or

$$T(x_i) - x_i - \big(T\big(x_j\big) - x_j\big) \neq z_b - z_a - x_i + x_j.$$

But if (i, j) is an edge in G, then the right-hand side of this equations equals zero when

$$z_b = \frac{x_i - x_j}{2}, z_a = -z_b,$$

and these are both in $\mathcal{Z}$. So we conclude that if i and j are neighbors in G, then

$$T(x_i) - x_i - \big(T\big(x_j\big) - x_j\big) \neq 0$$

or $\gamma(i) \neq \gamma(j)$. Thus $\gamma(i)$ satisfies the constraint in the definition of the l_2-chromatic number (i.e., Definition 3).

Finally, we observe that

$$\gamma^* \leq \frac{1}{n}\sum_{i=1}^{n} \gamma(i)^2 = \frac{1}{n}\sum_{i=1}^{n} (T(x_i) - x_i)^2,$$

and because the second-stage cost is zero, this is equal to the expected Witsenhausen cost, which is at most B by assumption. ∎

Note that Lemma 2 did not use that the support of X_0 is a Sidon set. The next lemma, which is just the converse of Lemma 2, will use this fact.

Lemma 3: Suppose $1 \leq B \leq n^2$. If the l_2-chromatic number of G is at most B, then the discrete Witsenhausen problem constructed in Definition 4 has a solution of cost at most B.

Before we give a proof of this lemma, we require the following fact.

Lemma 4: $(x_i - x_j)/2 \in \mathcal{Z}$ if and only if (i, j) is an edge in G.

Proof: One direction is one immediate from Definition 4. On the other hand, suppose $(x_i - x_j)/2 \in \mathcal{Z}$. This means there exist neighbors a, b in G such that

$$\frac{x_i - x_j}{2} = \frac{x_a - x_b}{2}$$

or

$$x_i + x_i + x_b + x_b = x_a + x_a + x_j + x_j$$

Since $x_a \neq x_b$ and $\{x_1, \ldots, x_n\}$ is a Sidon set of order four, this implies that $x_i = x_a, x_b = x_j$. Thus i and j are neighbors. ∎

Proof of Lemma 3: Paralleling the proof of Lemma 2, we define

$$T(x_i) = x_i + \gamma(i),$$

where γ is the coloring that achieves l_2-chromatic number at most B. For integers $x' \notin \{x_1, \ldots, x_n\}$, we can define $T(x')$ arbitrarily, as it does not affect the Witsenhausen cost. We will show that, with this choice, the second-stage cost is zero. Once this is shown, the proof will be complete as the l_2-chromatic number $(1/n)\sum_i \gamma^2(i)$ is just the transportation cost.

To argue that the second stage cost is zero, we proceed by contradiction. The second stage cost is not zero if and only if there exist $x_i, x_j \in \mathcal{X}, z_a, z_b \in \mathcal{Z}$ with $T(x_i) \neq T(x_j)$ such that

$$T(x_i) + z_a = T\big(x_j\big) + z_b \tag{4}$$

But, as in Lemma 2, we cannot have $x_i \neq x_j$ with $T(x_i) = T(x_j)$; indeed, by the same argument as Lemma 2, this implies that $|T(x_i) - x_i| > 2n^{1.5}$, which now contradicts the fact that γ achieves l_2-chromatic number at most $B \leq n^2$. So the second stage cost is zero if and only if there exist $x_i, x_j \in \mathcal{X}, x_i \neq x_j, z_a, z_b \in \mathcal{Z}$ such that Eq. (4) is satisfied. Now observe we can write Eq. (4)

$$x_i + \gamma(i) + z_a = x_j + \gamma(j) + z_b$$

or

$$x_i + z_a - x_j - z_b = \gamma(j) - \gamma(i). \tag{5}$$

Now the way Z was constructed in Definition 4 means that there exist neighbors c, d and neighbors e, f such that

$$z_a = \frac{x_c - x_d}{2}, \quad z_b = \frac{x_e - x_f}{2}.$$

Plugging this into Eq. (5) and doubling both sides,

$$2x_i + x_c - x_d - 2x_j - x_e + x_f = 2\big[\gamma(j) - \gamma(i)\big]$$

or

$$\big(x_i + x_i + x_c + x_f\big) - \big(x_j + x_j + x_d + x_e\big) = 2(\gamma(j) - \gamma(i)) \tag{6}$$

Now we consider two possibilities both of which lead to a contradiction. The left-hand side of Eq. (6) is either zero or nonzero.

If it is zero, then since $x_i \neq x_j$, and $\{x_1, \ldots, x_n\}$ being a Sidon set of order 4, we must have

$$x_i = x_d = x_e \quad \text{and} \quad x_j = x_c = x_f.$$

But this means that $(x_j - x_i)/2 = (x_c - x_d)/2 \in \mathcal{Z}$ so that by Lemma 3 we have that i and j are neighbors. But since the left-hand side of Eq. (6) is zero, we have that $\gamma(j) = \gamma(i)$ for a pair of neighbors i, j, a contradiction.

On the other hand, if the left-hand side of Eq. (6) is nonzero, then, since every x_i is a multiple of l by construction (recall Definition 4), the same left-hand side must have absolute value at least l. It follows that

$$|\gamma(j) - \gamma(i)| \geq \frac{l}{2} > 2n^{1.5},$$

where the strict inequality used the definition of l. Thus at least one of $|\gamma(i)|, |\gamma(j)|$ is strictly bigger than $n^{1.5}$. But this contradicts that the l_2-chromatic number is at most $B \leq n^2$. This concludes the proof. ∎

We now turn to an analysis of the l_2-chromatic number. We begin with a lemma which shows that the l_2-chromatic number is not very far from the ordinary chromatic number. Recall that we use the notation $\kappa(G)$ for the ordinary chromatic number of G.

Lemma 5:

$$\kappa^2 \geq \gamma^* \geq \frac{1}{n}\frac{(\kappa-2)^3}{12}.$$

Proof: For the first inequality, simply consider taking $\gamma(i)$ to be the color of vertex i, represented by an integer in the set $\{1, \ldots, \kappa\}$, using a coloring that minimizes the number of colors used.

For the second inequality, consider the optimal γ in the definition of l_2-coloring. Let us translate the γ so that the smallest interval I containing its range is symmetric about the origin, i.e., it equals either $[-a, a]$ or $[-a, a+1]$. Observe that every element in I is used, i.e., every element in I equals $\gamma(i)$ for some i, for else it would be possible to obtain a γ with smaller l_2 chromatic number. This implies that

$$\gamma^* \geq 2\frac{1}{n}\left(1^2 + \cdots + a^2\right) \geq \frac{2}{3}\frac{a^3}{n}.$$

On the other hand, the chromatic number is at most $2(a+1)$. Thus

$$\kappa \leq 2a + 2 \leq 2\left((3/2)n\gamma^*\right)^{1/3} + 2$$

or

$$(\kappa - 2)^3 \leq 12n\gamma^*,$$

which is a rearrangement of the second inequality. ∎

Lemma 6: Unless $P = NP$, for any $\epsilon > 0$, there exists no polynomial time algorithm which, given an undirected graph on n vertices, returns a number between γ^* and $n^{2-\epsilon}\gamma^*$.

Proof: It is possible to define the notion of a *fractional chromatic number* of a graph G, denoted by $\chi_f(G)$. We avoid giving a definition here because we only need to use the following two facts about it:

- In [14, Th. 1.2], it was shown that, for any $\epsilon > 0$, it is NP-hard to distinguish between graphs G on n vertices with fractional chromatic number of n^{ϵ} from graphs with fractional chromatic number of $n^{1-\epsilon}$.
- In [15], it was shown that the fractional chromatic number is a logarithmic approximation to the chromatic number, i.e.,

$$\frac{\kappa(G)}{1+\log n} \leq \chi_f(G) \leq \kappa(G),$$

where, recall, $\kappa(G)$ is the ordinary chromatic number; for more details, see the discussion in [16, Sec. 3.3].

As remarked in [16], these two facts imply that it is NP-hard to distinguish between graphs of chromatic number $n^{\epsilon}(1+\log n)$ and graphs with chromatic number $n^{1-\epsilon}$.

Algorithm 1 Approximation Algorithm for the Discrete Witsenhausen Problem

1: Input: distributions of X_0, Z
2: **for** $k = 0, \ldots, n$ **do**
3: Set T^k be a map that map that fixes $x_1, \ldots, x_k$ and maps $x_{k+1}, \ldots, x_n$ to values ensuring there are no collisions except between $x_1, \ldots, x_k$.
4: Choose δ^k to be the optimal second-stage map given T^k.
5: **end for**
6: Choose the pair among $(T^k, \delta^k), k = 1, \ldots, n$ with lowest Witsenhausen cost.

Applying Lemma 5, it follows that it is NP-hard to distinguish between graphs with $\gamma^* \leq n^{2\epsilon}(1+\log n)^2$ and graphs with $\gamma^* \geq \frac{1}{12n}(n^{1-\epsilon}-2)^3$. We conclude that, for any $\epsilon > 0$, it is NP-hard to approximate γ^* within a multiplicative factor of less than

$$\frac{(n^{1-\epsilon}-2)^3}{12n^{1+2\epsilon}(1+\log n)^2}.$$

Because this quantity can be lower bounded by $n^{2-O(\epsilon)}$, this completes the proof. ∎

Finally, we are now able to provide a proof of our main result.

Proof of Theorem 1: Consider a graph G with l_2-chromatic number of B. Since every vertex can be colored by a different color, we have that $B \leq n^2$. Consider the discrete Witsenhausen problem constructed in Definition 4: putting together Lemma 2 and Lemma 3, we obtain that its optimal solution has cost B. Now observing that by Lemma 6, it is NP-hard to approximate B to within a multiplicative factor of $n^{2-\epsilon}$ completes the proof. ∎

IV. Proof of Theorem 2

We now describe an algorithm for the Witsenhausen problem whose approximation ratio is polynomial in $|\mathcal{X}|$ and $|\mathcal{Z}|$. We begin with an informal discussion intended to motivate our approach. Parallelling our arguments in the previous section, we'll adopt the convention of saying that i and j "collide" if

$$T(x_i) + z_a = T\left(x_j\right) + z_b, \tag{7}$$

for some $z_a, z_b \in \mathcal{Z}$.

Our approach is simple: we "interpolate" between the optimal solution when $K = 0$ (which results in $T(x_i) = x_i$) and $K \to +\infty$ (which results in a T that avoids any collisions) by fixing the k elements in $\mathcal{X}$ with the highest probabilities, and moving all the other entries in $\mathcal{X}$ to avoid collisions. We do this for all $k = 1, \ldots, n$ where $n = |\mathcal{X}|$ and choose the best result.

We outline the approach in the algorithm box below, where we use the convention that p_i is the probability of $X_0 = x_i$ and

$$p_1 \geq p_2 \geq \cdots \geq p_n.$$

It is easy to see that this is a polynomial-time algorithm. Indeed, step 6 can easily be done in polynomial time: the

cost of each pair T^i, δ^i is a sum over $|\mathcal{X}||\mathcal{Z}|$ values. Second, step 4 can also be done without difficulty, since the selection of the best second-stage map given the transportation map is an ordinary least-squares estimation problem. The following lemma discusses how to do step 3 and implicitly gives an upper bound on the transportation cost of the T^i chosen in that step.

Lemma 7: Step 3 can be done in polynomial time with $|T^k(x_j) - x_j| \leq |\mathcal{X}||\mathcal{Z}|^2$ for all j.

Proof: Starting with $j = k+1$, we sequntially set $T(x_j)$ to be the closest value to x_j that does not yield a collision; when we have set $T(x_n)$, we are done. When we consider x_m, looking at Eq. (7), we have to avoid

$$T(x_m) = T\big(x_j\big) + z_b - z_a, \quad j < m, \ z_a, z_b \in \mathcal{Z},$$

which rules out at most $(m-1)|\mathcal{Z}|^2$ different values. It follows that we can always assign $T(x_m)$ so that $|T(x_m) - x_m| \leq |\mathcal{X}||\mathcal{Z}|^2$. Moreover, each step of this procedure requires examining at most $|\mathcal{X}||\mathcal{Z}|^2$ possibilities, and the number of steps is at most $|\mathcal{X}|$, so this procedure is polynomial time. ■

We can now proceed to the proof of Theorem 2. Our first step is to introduce some notation. We let $\Phi_1(p_{X_0}, p_Z, T, \delta)$ to be the first-stage (transporation) cost when X_0, Z, T, δ are the random variables and maps in the discrete Witsenhausen problem. Likewise, we will use $\Phi_2(p_{X_0}, p_Z, T, \delta)$ to denote the second-stage cost. Occasionally, we will omit to write the δ in this notation, and it should be understood that δ is then selected to be the optimal choice for the given T.

Proof of Theorem 2: We claim that Algorithm 1 with the selection procedure of Lemma 7 returns a solution with cost $|\mathcal{X}|^3|\mathcal{Z}|^4\Phi^*$ where Φ^* is the optimal Witsenhausen cost.

Indeed, consider the optimal strategy T^*, δ^*. Let l be the smallest index such that $T^*(x_l) \neq x_l$ (we can assume such an index exists, because otherwise Algorithm 1 finds the optimal solution when $k = n$ and there is nothing to prove). The transport cost incurred by T^* is at least p_l.

Now consider the (T^k, δ^k) when $k = l$. The transport cost incurred by T^l is upper bounded by $(|\mathcal{X}|p_l)(|\mathcal{X}||\mathcal{Z}|^2)^2$ because the probability of not landing at a fixed point is at most $|\mathcal{X}|p_l$, in which case one moves by at most $|\mathcal{X}||\mathcal{Z}|^2$ as a consequence of Lemma 7. Thus the transport cost incurred by T^l is at most $p_l|\mathcal{X}|^3|\mathcal{Z}|^4$. Putting the last two paragraphs together,

$$\Phi_1\Big(p_{X_0}, p_Z, T^l, \delta^l\Big) \leq |\mathcal{X}|^3|\mathcal{Z}|^4\Phi_1^*\big(p_{X_0}, p_Z\big) \tag{8}$$

We now consider the second-stage cost of T^l, δ^l. By construction whenever one of $(x_{l+1}, \ldots, x_n)$ is generated, the second-stage cost is zero. Defining p' to be the distribution proportional to $(p_1, p_2, \ldots, p_{l-1})$, this means that

$$\Phi_2\Big(p_{X_0}, p_Z, T^l, \delta^l\Big) = (p_1 + \cdots p_{l-1})\Phi_2\big(p', p_Z, I\big). \tag{9}$$

where we use I for the identity map and we used that T^l fixes $x_1, \ldots, x_{l-1}$.

On the other hand, consider the second stage cost under T^*, δ^*. Let A be the event that $X_0 \in \{x_1, \ldots, x_{l-1}\}$. The second-stage cost cannot be increased if the first agent transmits to the second agent whether A has occurred or not. Thus

$$\Phi_2\big(p_{X_0}, p_Z, T^*, \delta^*\big) \geq (p_1 + \cdots p_{l-1})\Phi_2\big(p', p_Z, I\big). \tag{10}$$

Finally, comparing Eq. (9) and Eq. (10) we obtain $\Phi_2(p_{X_0}, p_Z, T^*, \delta^*) \geq \Phi_2(p_{X_0}, p_Z, T^l, \delta^l)$. Putting this together with Eq. (8) completes the proof. ■

ACKNOWLEDGMENT

The author would like to thank Dr. S. Kopparty and Dr. A. Bhangale for suggesting the reduction between the l_2-chromatic number and the ordinary chromatic number. The author would also like to acknowledge Dr. M. Agarwal for multiple discussions of this problem.

REFERENCES

[1] C. H. Papadimitriou and J. N. Tsitsiklis, "Intractable problems in control theory," *SIAM J Control Optim.*, vol. 24, no. 4, pp. 639–654, 1986.
[2] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J Control*, vol. 6, no. 1, pp. 131–147, 1968.
[3] S. Yüksel and T. Basar, *Stochastic Networked Control Systems*. New York, NY, USA: Birkhäuser, 2013.
[4] T. Basar, "Variations on the theme of the Witsenhausen counterexample," in *Proc. 47th IEEE Conf. Decis. Control (CDC)*, 2008, pp. 1614–1619.
[5] C. A. Uribe, T. Keviczky, and J. H. van Schuppen, "Computing optimal control laws for finite stochastic systems with non-classical information patterns," in *Proc. IEEE Amer. Control Conf. (ACC)*, 2014, pp. 5742–5747.
[6] A. Gupta, S. Yüksel, T. Başar, and C. Langbort, "On the existence of optimal policies for a class of static and sequential dynamic teams," *SIAM J Control Optim.*, vol. 53, no. 3, pp. 1681–1712, 2015.
[7] A. A. Kulkarni and T. P. Coleman, "An optimizer's approach to stochastic control problems with nonclassical information structures," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 937–949, Apr. 2015.
[8] S. T. Jose and A. A. Kulkarni, "A linear programming relaxation for stochastic control problems with non-classical information patterns," in *Proc. IEEE 54th Annu. Decis. Control Conf. (CDC)*, 2015, pp. 5743–5748.
[9] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1644–1658, Jul. 2013.
[10] Y. Ho and T. Chang, "Another look at the nonclassical information structure problem," *IEEE Trans. Autom. Control*, vol. 25, no. 3, pp. 537–540, Jun. 1980.
[11] N. Saldi, S. Yüksel, and T. Linder, "Finite model approximations and asymptotic optimality of quantized policies in decentralized stochastic control," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2360–2373, May 2017.
[12] M. Cygan, "Improved approximation for 3-dimensional matching via bounded pathwidth local search," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci. (FOCS)*, 2013, pp. 509–518.
[13] P. Grover, A. Sahai, and S. Y. Park, "The finite-dimensional Witsenhausen counterexample," in *Proc. 7th Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw. (WiOPT)*, 2009, pp. 1–10.
[14] D. Zuckerman, "Linear degree extractors and the inapproximability of max clique and chromatic number," in *Proc. 38th Annu. ACM Symp. Theory Comput.*, 2006, pp. 681–690.
[15] L. Lovász, "On the ratio of optimal integral and fractional covers," *Disc. Math.*, vol. 13, no. 4, pp. 383–390, 1975.
[16] U. Feige and J. Kilian, "Zero knowledge and the chromatic number," *J. Comput. Syst. Sci.*, vol. 57, no. 2, pp. 187–199, 1998.

Deterministic and Randomized Actuator Scheduling With Guaranteed Performance Bounds

Milad Siami, *Member, IEEE*, Alexander Olshevsky, *Member, IEEE*, and Ali Jadbabaie, *Fellow, IEEE*

***Abstract*—In this article, we investigate the problem of actuator selection for linear dynamical systems. We develop a framework to design a sparse actuator schedule for a given large-scale linear system with guaranteed performance bounds using deterministic polynomial-time and randomized approximately linear-time algorithms. First, we introduce systemic controllability metrics for linear dynamical systems that are monotone and homogeneous with respect to the controllability Gramian. We show that several popular and widely used optimization criteria in the literature belong to this class of controllability metrics. Our main result is to provide a polynomial-time actuator schedule that on average selects only a constant number of actuators at each time step, independent of the dimension, to furnish a guaranteed approximation of the controllability metrics in comparison to when all actuators are in use. Our results naturally apply to the dual problem of sensor selection, in which we provide a guaranteed approximation to the observability Gramian. We illustrate the effectiveness of our theoretical findings via several numerical simulations using benchmark examples.**

***Index Terms*—Approximation algorithm, complexity theory, controllability, dynamic scheduling, linear dynamical systems, sparse sensor and actuator selections.**

I. Introduction

OVER the past few years, controllability and observability properties of complex dynamical networks have been subjects of intense study in the controls community [1]–[12]. This interest stems from the need to steer or observe the state of large-scale, networked systems such as the power grids [13], social networks, biological and genetic regulatory networks [14]–[16], and traffic networks [17]. While the classical notion of controllability and observability, introduced by Kalman in [18] is quite well understood, the dependence of various measures of controllability or observability on number and location of sensors and actuators in linear systems have been subject of study for nearly five decades [19]. Often times, there is a need to steer or estimate the state of a large-scale, networked control system with as few actuators/sensors as possible, due to issues related to cost and energy depletion. The desire to perform control/estimation using a sparse set of actuators/sensors spans various application domains, ranging from infrastructure networks (e.g., water and power networks) to multirobot systems and the study of the human connectome. For example, energy conservation through efficient utilization of sensors and actuators can help extend the duration of battery life in networks of mobile sensors and multiagent robotic networks; estimating the whole state of the power grid using fewer measurement units will help reduce the cost of monitoring the network for systemic failures, etc.

It is, therefore, desirable to have a limited number of sensors and actuators without compromising the control or estimation performance too much. Unfortunately, as the recent works in [1] and [6] have shown, the problem of finding a sparse set of input variables such that the resulting system is controllable is NP-hard. Even the presumably easier problem of approximating the minimum number better than a constant multiplicative factor of $\log n$ is also NP-hard. Other results in the literature have studied network controllability by exploring approximation algorithms for the closely related subset selection problem [1], [11], [12]. More recently, some of the authors showed that even the problem of finding a sparse set of actuators to guarantee reachability of a particular state is hard and even hard to approximate [20].

Previous studies have been mainly focused on solving the optimal sensor/actuator placement problem using the greedy heuristic, as approximations of the corresponding sparse-subset selection problem. While these results attempt to find approximation algorithms for finding the best sparse subset, our focus in this article is to gain new fundamental insights into approximating various controllability metrics compared to the case when all possible actuators are chosen. Specifically, we are interested in actuator/sensor schedules that select a small number of actuators/sensors so as to save the energy while ensuring a suitable level of controllability (observability) performance

Manuscript received December 4, 2019; revised April 6, 2020; accepted May 26, 2020. Date of publication June 9, 2020; date of current version March 29, 2021. This work was supported by the Vannevar Bush Fellowship from the Office of Secretary of Defense. Recommended by Associate Editor F. Wirth. *(Corresponding author: Milad Siami.)*

Milad Siami is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: m.siami@northeastern.edu).

Alexander Olshevsky is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: alexols@bu.edu).

Ali Jadbabaie is with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: jadbabai@mit.edu).

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

Digital Object Identifier 10.1109/TAC.2020.3000976

for the entire network. Due to energy efficiency, we may want to reduce the number of active actuator/sensors at each time. At the same time, we would like to have a performance that closely resembles that of the original system, when all available sensor/actuators are active.

We investigate sparse sensor and actuator selection as particular instances, where discrete geometric structures can be utilized to study network controllability and observability problems (cf. [20]–[22]). A key observation is the close connection between this problem and some classical problems in statistics such as outlier detection, active learning, and optimal experimental design. In recent years, there has been a renewed interest in optimal experiment design, which has a long history going back at least 65 years [23], [24].

One of our main contributions is to show that the time-varying actuator selection problem, which goes back to a paper by Athans [19], can be solved via random sampling. We propose an alternative to submodularity-based methods and instead use recent advances in theoretical computer science to develop scalable algorithms for sparsifying control inputs. Current approaches based on polynomial time relaxations of the subset selection problem require an extra multiplicative factor of $\log n$ sensors/actuators times the minimal number in order to just maintain controllability/observability. Using these recent advances [24]–[30], we show that by carefully designing a scheduling strategy, one can choose on average a constant number of sensors and actuators at each time, to approximate the controllability/observability metrics of the system when all sensors and actuators are in use.

Potential application domains. One potential application can be considered as widearea oscillation damping control using high voltage dc (HVdc) lines (e.g., [31]–[33]). HVdc systems are increasingly being installed in power grids all around the globe. This trend is expected to continue with recent advancements in power electronics technology, energy harvesting, and usage of renewable energy [31]. In this setup, it seems quite compelling to examine approaches that can support sparse HVdc lines (i.e., actuators) scheduling to improve the controllability of the power grid in order to account for issues related to cost, energy depletion, and the limitations in directly accessing actuators, especially in large networks (cf. Example 2 in Section VII). Moreover, note that the dual problem of actuator scheduling for control is sensor scheduling for estimation. In the present case, our sparse sensor schedule setup is equal to reducing the number of measurements for data reduction, and the observability Gramian-based measures show how well one can estimate the state of the system [34].

Another potential application is disease spread estimation in networks where testing resources are scarce. There are several models for the spread of infections (see [35] and references therein). Formally, let us consider a network with n nodes. Each node represents a city and has a nonnegative scalar state $x_i(k)$ associated with it, which indicates the prevalence of an infection in that node. Since $x_i(t)$ is the *proportion* of the population at node i infected at time t, we assume that the $x_i(t)$ are close to zero (for example, this is valid for the recent COVID-19 pandemic; even though there are a substantial number of infected people still the proportion of the population that is infected is small as of early-April, 2020).[1] It, therefore, makes sense to linearize the epidemic models around the zero state.

After linearization, the states evolve according to an autonomous linear differential equation $x(k+1) = Ax(k)$; in all epidemic models, the OFF-diagonal entries a_{ij} of the state matrix A indicates the unitized transmission rate of the infection from city j to city i, while the diagonal entries can be positive or negative, reflecting the possibility of either local spread or recovery.

We assume $y(t) = C(t)x + w(t)$, where $C(t)$ is a "subset" of the identity matrix (because one can measure the prevalence of the infection in a node by randomly testing from the population at that node) and $w(t)$ is noise. The sensor scheduling problem in this context amounts estimating the state with as small a variance as possible, while the measurements have to be done over a certain time-horizon and are bounded in number due to scarce resources.[2]

Some of our results appeared earlier in the conference version of this article [36], [37]; however, their proofs are presented here for the first time. The manuscript also contains several new results, remarks, numerical examples, and proofs.

II. Preliminaries and Definitions

A. Mathematical Notations

Throughout this article, discrete time index is denoted by k. The sets of real (integer), nonnegative real (integer), and positive real (integer) numbers are represented by $\mathbb{R}$ ($\mathbb{Z}$), $\mathbb{R}_+$ ($\mathbb{Z}_+$), and $\mathbb{R}_{++}$ ($\mathbb{Z}_{++}$), respectively. The set of natural numbers $\{i \in \mathbb{Z}_{++} : i \leq n\}$ is denoted by $[n]$. The cardinality of a set σ is denoted by $\mathrm{card}(\sigma)$. Capital letters, such as A or B, stand for real-valued matrices. For a square matrix X, $\det(X)$, and $\mathrm{trace}(X)$ refer to the determinant and the summation of ON-diagonal elements of X, respectively. $\mathbb{S}_+^n$ is the positive definite cone of n-by-n matrices. The n-by-n identity matrix is denoted by I. Notation $A \preceq B$ is equivalent to matrix $B - A$ being positive semidefinite. The transpose of matrix A is denoted by $A^\top$. The rank, kernel and image of matrix A are referred to by $\mathrm{rank}(A)$, $\ker(A)$, and $\mathrm{Im}(A)$, respectively. The Moore–Penrose pseudoinverse of matrix A is denoted by $A^\dagger$. The ceiling function of $x \in \mathbb{R}$ is denoted by $\lceil x \rceil$, where it returns the least integer greater than or equal to x.

B. Linear Systems and Controllability

We start with the canonical linear discrete-time, time-invariant dynamics

$$x(k+1) = A\,x(k) + B\,u(k)$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $k \in \mathbb{Z}_+$. The state matrix A describes the underlying structure of the system and the interaction strength between the agents, and matrix B represents how

[1] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html

[2] From now on we will focus the article on the actuator selection problem. The dual notion of sensor selection follows similar ideas.

the control input enters the system. Equivalently, the dynamics can be written as

$$x(k+1) = A\,x(k) + \sum_{i\in[m]} b_i\,u_i(k) \tag{1}$$

where b_i's are columns of matrix $B \in \mathbb{R}^{n\times m}$. Then, the controllability matrix at time t is given by

$$\mathcal{C}(t) = \begin{bmatrix} B & AB & A^2B & \cdots & A^{t-1}B \end{bmatrix}. \tag{2}$$

In this article, we assume that $t > 0$ is the time horizon to control (also known as the time-to-control). It is well-known that from a numerical standpoint it is better to characterize controllability in terms of the Gramian matrix at time t defined as follows:

$$\mathcal{W}(t) = \sum_{i=0}^{t-1} A^i BB^\top (A^i)^\top = \mathcal{C}(t)\,\mathcal{C}^\top(t). \tag{3}$$

When looking at time-varying input/actuator schedules, we will consider the following linear system with time-varying input matrix $\mathcal{B}(.)$

$$x(k+1) = A\,x(k) + \mathcal{B}(k)\,u(k). \tag{4}$$

For the abovementioned system, the controllability and Gramian matrices at time step t are defined as

$$\mathcal{C}_*(t) = \begin{bmatrix} \mathcal{B}(t-1) & A\mathcal{B}(t-2) & A^2\mathcal{B}(t-3) \cdots A^{t-1}\mathcal{B}(0) \end{bmatrix}$$

and

$$\begin{aligned} \mathcal{W}_*(t) &= \sum_{i=0}^{t-1} A^i \mathcal{B}(t-i-1)\mathcal{B}^\top(t-i-1)(A^i)^\top \\ &= \mathcal{C}_*(t)\,\mathcal{C}_*^\top(t) \end{aligned} \tag{5}$$

respectively.

Assumption 1: Throughout this article, we assume that the system (1) is controllable (i.e., the controllability matrix has full row rank and the Gramian is positive definite). However, all results presented in this article can be modified/extended to uncontrollable systems.

C. Matrix Reconstruction and Sparsification

The key idea throughout the article is to approximate the time-t controllability Gramian as a sparse sum of rank-1 matrices, while controlling the approximation error. To this end, we present a key lemma from the sparsification literature and state the necessary modification. We, then, use this result later in our deterministic algorithm to find a sparse actuator schedule.

Lemma 1 (Dual Set Spectral Sparsification [38]): Let $V = \{v_1, \ldots, v_t\}$ and $U = \{u_1, \ldots, u_t\}$ be two equal cardinality decompositions of identity matrices (i.e., $\sum_{i=1}^t v_i v_i^\top = I_n$ and $\sum_{i=1}^t u_i u_i^\top = I_\ell$) where $v_i \in \mathbb{R}^n$ $(n < t)$ and $u_i \in \mathbb{R}^\ell$ $(\ell \le t)$. Given an integer κ with $n < \kappa \le t$, Algorithm 1 computes a set of weights $c_i \ge 0$ where $i \in [t]$, such that

$$\lambda_{\min}\left(\sum_{i=1}^t c_i v_i v_i^\top\right) \ge \left(1 - \sqrt{\frac{n}{\kappa}}\right)^2$$

$$\lambda_{\max}\left(\sum_{i=1}^t c_i u_i u_i^\top\right) \le \left(1 + \sqrt{\frac{\ell}{\kappa}}\right)^2$$

and

$$\operatorname{card}\{i : c_i > 0, i \in [t]\} \le \kappa.$$

Due to space limitations, we refer the interested readers to [38] for more details on Algorithm 1. However, roughly speaking, Algorithm 1 is based on choosing vectors in a greedy fashion that satisfy a set of desired properties at each step, leading to bounds on eigenvalues. In Algorithm 1, lower and upper barriers or potential functions are defined as follows:

$$\underline{\phi}(\underline{\mu}, \underline{\mathcal{A}}) = \sum_{i=1}^n \frac{1}{\lambda_i(\underline{\mathcal{A}}) - \underline{\mu}} \tag{6}$$

and

$$\bar{\phi}(\bar{\mu}, \bar{\mathcal{A}}) = \sum_{i=1}^\ell \frac{1}{\bar{\mu} - \lambda_i(\bar{\mathcal{A}})} \tag{7}$$

respectively. These potential functions quantify how far the eigenvalues of $\underline{\mathcal{A}}$ and $\bar{\mathcal{A}}$ are from the barriers $\underline{\mu}$ and $\bar{\mu}$. These potential functions become unbounded as any eigenvalue nears the barriers.[3] We control the maximum eigenvalue of $\bar{\mathcal{A}}$ using an upper barrier $\bar{\mu}$ and the minimum eigenvalue of $\underline{\mathcal{A}}$ using a lower barrier $\underline{\mu}$. Two parameters $\mathfrak{L}$ and $\mathfrak{U}$ are defined as follows:

$$\begin{aligned} &\mathfrak{L}(v, \underline{\delta}, \underline{\mathcal{A}}, \underline{\mu}) \\ &= \frac{v^\top \left(\underline{\mathcal{A}} - (\underline{\mu} + \underline{\delta})I_n\right)^{-2} v}{\underline{\phi}(\underline{\mu} + \underline{\delta}, \underline{\mathcal{A}}) - \underline{\phi}(\underline{\mu}, \underline{\mathcal{A}})} - v^\top \left(\underline{\mathcal{A}} - (\underline{\mu} + \underline{\delta})I_n\right)^{-1} v \end{aligned}$$

and

$$\begin{aligned} &\mathfrak{U}(u, \bar{\delta}, \bar{\mathcal{A}}, \bar{\mu}) \\ &= \frac{u^\top((\bar{\mu} + \bar{\delta})I_\ell - \bar{\mathcal{A}})^{-2} u}{\bar{\phi}(\bar{\mu}, \bar{\mathcal{A}}) - \bar{\phi}(\bar{\mu} + \bar{\delta}, \bar{\mathcal{A}})} + u^\top \left((\bar{\mu} + \bar{\delta})I_\ell - \bar{\mathcal{A}}\right)^{-1} u. \end{aligned}$$

The Sherman–Morrison–Woodbury formula inspires the structure of the abovementioned quantities for more details on the barrier method (cf. [40, Sec. 1.2]). These potential functions (6) and (7) are chosen to guide the selection of vectors and scalings at each timestep τ and to ensure steady progress of the algorithm. Small values of these potentials indicate that the eigenvalues of $\bar{\mathcal{A}}$ and $\underline{\mathcal{A}}$ do not concentrate near $\bar{\mu}$ and $\underline{\mu}$, respectively. In Algorithm 1, at each iteration, we increase the upper barrier $\bar{\mu}$ by a fixed constant $\bar{\delta}$ and the lower barrier $\bar{\mu}$ by another fixed constant $\underline{\delta}$. It can be shown that as long as the potentials remain bounded, there must exist (at every step τ) a choice of an index j and weight c_j so that the addition

[3] These potentials are equal to constant multiples of the Stieltjes transform of $\underline{\mathcal{A}}$ and $\bar{\mathcal{A}}$ evaluated at $\underline{\mu}$ and $\bar{\mu}$, respectively [39].

Algorithm 1: A Deterministic Dual Set Spectral Sparsification $\mathrm{DualSet}(V, U, \kappa)$.

Input : $V = [v_1, \ldots, v_t]$, with $VV^\top = I_n$
$U = [u_1, \ldots, u_t]$, with $UU^\top = I_\ell$
$\kappa \in \mathbb{Z}_+$, with $n < \kappa \leq t$

Output: $c = [c_1, c_2, \ldots, c_t] \in \mathbb{R}_+^{1\times t}$ with $\|c\|_0 \leq \kappa$

1 Set $c(0) = 0_{t\times 1}$, $\underline{\mathcal{A}}(0) = 0_{n\times n}$, $\bar{\mathcal{A}}(0) = 0_{\ell\times\ell}$, $\underline{\delta} = 1$, $\bar{\delta} = \frac{1+\sqrt{\frac{\ell}{\kappa}}}{1-\sqrt{\frac{n}{\kappa}}}$
2 **for** $\tau = 0 : \kappa - 1$ **do**
3 $\quad \underline{\mu}(\tau) = \tau - \sqrt{\kappa n}$
4 $\quad \bar{\mu}(\tau) = \bar{\delta}\left(\tau + \sqrt{\kappa\ell}\right)$
5 $\quad$ Find an index j such that

$$\mathfrak{U}(u_j, \bar{\delta}, \bar{\mathcal{A}}(\tau), \bar{\mu}(\tau)) \leq \mathfrak{L}(v_j, \underline{\delta}, \underline{\mathcal{A}}(\tau), \underline{\mu}(\tau))$$

6 $\quad$ Set $\Delta = 2\big(\mathfrak{U}(u_j, \bar{\delta}, \bar{\mathcal{A}}(\tau), \bar{\mu}(\tau)) + \mathfrak{L}(v_j, \underline{\delta}, \underline{\mathcal{A}}(\tau), \underline{\mu}(\tau))\big)^{-1}$
7 $\quad$ Update the j-th component of $c(\tau)$:

$$c(\tau+1) = c(\tau) + \Delta \mathrm{e}_j,$$

8 $\quad \underline{\mathcal{A}}(\tau+1) = \underline{\mathcal{A}}(\tau) + \Delta v_j v_j^\top$
9 $\quad \bar{\mathcal{A}}(\tau+1) = \bar{\mathcal{A}}(\tau) + \Delta u_j u_j^\top$
10 **end**
11 **return** $c = \kappa^{-1}\left(1 - \sqrt{\frac{n}{\kappa}}\right) c(\kappa)$

of associated rank-1 matrices to $\bar{\mathcal{A}}$ and $\underline{\mathcal{A}}$, and the increments of barriers do not increase either potential and keep all the eigenvalues of the updated matrix between the barriers (see Algorithm 1). Repeating these steps ensures steady growth of all the eigenvalues and yields the desired result.

This algorithm is a generalization of an algorithm from [27], which is deterministic and at most needs $\mathcal{O}(\kappa t(n^2 + \ell^2))$. Furthermore, the algorithm needs $\mathcal{O}(\kappa t n^2)$ operations if U contains the standard basis of $\mathbb{R}^t$; we refer the reader to [38] for more details.

Remark 1: We modify the fifth line of Algorithm 1; at each step, we choose an index j that maximizes

$$\mathfrak{L}(v_j, \underline{\delta}, \underline{\mathcal{A}}(\tau), \underline{\mu}(\tau)) - \mathfrak{U}(u_j, \bar{\delta}, \bar{\mathcal{A}}(\tau), \bar{\mu}(\tau)) \tag{8}$$

instead of only finding an index j such that

$$\mathfrak{U}(u_j, \bar{\delta}, \bar{\mathcal{A}}(\tau), \bar{\mu}(\tau)) \leq \mathfrak{L}(v_j, \underline{\delta}, \underline{\mathcal{A}}(\tau), \underline{\mu}(\tau)). \tag{9}$$

We should note that if an index j maximizes (8), then it will satisfy (9). Therefore, Lemma 1 still holds for the modified algorithm, and hence, the theoretical bounds are valid. Based on our simulations, we observe that this modification can help to improve Algorithm 1 by producing smaller ratio $\lambda_{\max}(\sum_{i=1}^t c_i v_i v_i^\top)/\lambda_{\min}(\sum_{i=1}^t c_i v_i v_i^\top)$ (in Section V, we will see that this quantity is closely related to approximation factor ϵ). We denote the application of the algorithm to V and U by

$$[c_1, c_2, \ldots, c_t] = \mathrm{DualSet}^*(V, U, \kappa).$$

We now recall the concentration lemma of Rudelson–Vershynin [28] as follows. We are going to use this result in the proof of Theorem 2.

Lemma 2 (Th. 3.1[28]): Let $y \in \mathbb{R}^p$ be a random vector such that $\|y\| \leq b$ almost surely and $\|\mathbf{E}\, yy^\top\|_2 \leq 1$. Let $y_1, \ldots, y_n$ be i.i.d. copies of y. Then

$$\mathbf{E}\left\|\frac{1}{n}\sum_{i=1}^n y_i y_i^\top - \mathbf{E}\, yy^\top\right\|_2 \leq \min\left(1, cb\sqrt{\frac{\log n}{n}}\right) \tag{10}$$

where $c > 0$ is some universal constant.

In the following section, we show how various controllability measures can be approximated by selecting a sparse set of actuators via the abovementioned algorithm.

TABLE I
SOME IMPORTANT EXAMPLES OF SYSTEMIC CONTROLLABILITY METRICS

Optimality-criteria	Systemic Controllability Measure	Matrix Operator Form
A-optimality	Average control energy	$\mathrm{trace}\left(\mathcal{W}^{-1}(t)\right)$
D-optimality	The volume of the ellipsoid	$(\det \mathcal{W}(t))^{-1/n}$
T-optimality	Inverse of the trace	$1/\mathrm{trace}(\mathcal{W}(t))$
E-optimality	Inverse of the minimum eigenvalue	$1/\lambda_{\min}(\mathcal{W}(t))$

III. SYSTEMIC CONTROLLABILITY METRICS

Similar to the *systemic* notions introduced in [41]–[43], we define various controllability metrics. These measures are real-valued operators defined on the set of all linear dynamical systems governed by (4) and quantify various measures of the required control energy. All of the metrics depend on the controllability Gramian matrix of the system, which is a positive definite matrix. Therefore, one can define a systemic controllability performance measure as an operator on the set of Gramian matrices of all controllable systems with n states which we represent by $\mathbb{S}_+^n$.[4]

Definition 1 (Systemic Criteria): A controllability metric $\rho : \mathbb{S}_+^n \to \mathbb{R}$ is systemic if and only if

1) *Homogeneity:* For all $\kappa > 1$

$$\rho(\kappa A) = \kappa^{-1}\rho(A).$$

2) *Monotonicity:* If $B \preceq A$, then

$$\rho(A) \leq \rho(B).$$

For many popular choices of ρ, one can see that they satisfy the properties presented in Definition 1. Some of them are listed in Table I. We note that similar criteria have been developed [23], [24], [44] in the experiment design literature (cf. Table I). In what follows, we will make this statement formal.

Proposition 1: For given dynamics (4) with Gramian matrix $\mathcal{W}(t)$, the metrics presented in Table 1 are systemic controllability measures.

Proof: One can easily see that all these measures satisfy the homogeneity, and monotonicity properties in Definition 1 (cf. [43], [45]). ■

In the following section, we show how various measures can be approximated by selecting a sparse set of actuators.

[4]For any $X \in \mathbb{S}_+^n$ and given $t \in \mathbb{Z}_{++}$, there exists at least one controllable system with $\mathcal{W}(t) = X$ (e.g., $x(k+1) = X^{\frac{1}{2}}u(k)$), and for any controllable system, it is well known that the Gramian matrix is positive definite [see (5)]. Therefore, the set of Gramian matrices of all controllable systems with n states is equal to $\mathbb{S}_+^n$.

IV. Sparse Actuator Selection Problems

For a given linear system (1) with a general underlying structure, the actuator scheduling problem seeks to construct a schedule of the control inputs that keeps the number of active actuators much less than the original system such that the controllability matrices of the original and the new systems are similar in an appropriately defined sense. Specifically, given a canonical linear, time-invariant system (1) with m actuators and controllability Gramian matrix $\mathcal{W}(t)$ at time t, our goal is to find a sparse actuator schedule such that the resulting system with controllability Gramian $\mathcal{W}_s(t)$ is well-approximated, i.e.,

$$\left|\frac{\rho\left(\mathcal{W}(t)\right)-\rho\left(\mathcal{W}_s(t)\right)}{\rho\left(\mathcal{W}(t)\right)}\right| \leq \epsilon \tag{11}$$

where ρ is any systemic controllability metric that quantifies the difficulty of the control problem for example as a function of the required control energy, and $\epsilon \geq 0$ is the approximation factor. The systemic controllability metrics are defined based on the controllability Gramian, therefore "close" Gramian matrices result in approximately the same values. Our goal here is to answer the following questions.

1) What is the minimum number of actuators to be chosen to achieve a good approximation of the system with the full set of actuators utilized?
2) What is the relation between the number of selected actuators and performance/controllability loss?
3) Does a sparse approximation schedule exist with at most a constant number of active actuators at each time?
4) What is the time complexity of choosing the subset of actuators with guaranteed performance bounds?

In the rest of this article, we show how some fairly recent advances in theoretical computer science and the probabilistic method can be utilized to answer these questions. The probabilistic method is one of the most important tools of modern combinatorics, which was introduced by Erdös. The idea is that a deterministic solution is shown to exist by constructing a random candidate satisfying all the requirements of the problem with positive probability. Recently, Marcus *et al.* [29] introduced a new variant of the probabilistic method, which ends up solving the so-called Kadison–Singer (KS) conjecture. We use the solution approach to the KS conjecture together with a combination of tools from Sections V to find a sparse approximation of the actuator selection problem with algorithms that have favorable time-complexity.

V. A Weighted Sparse Actuator Schedule

As a starting point, we allow for scaling of the input signals at chosen inputs while keeping the input scaling bounded. The input scaling allows for an extra degree of freedom that could allow for choosing a sparser set of inputs. Given (1), we define a weighted actuator schedule by $\sigma = \{\sigma_k\}_{k=0}^{t-1}$ and scalings $s_i(k) \geq 0$, where $i \in [m]$, $k+1 \in [t]$, and $\sigma_k = \{i | s_i(k) > 0\} \subseteq [m]$. The resulting system with this schedule is

$$x(k+1) = A\,x(k) + \sum_{i\in\sigma_k} s_i(k)\,b_i\,u_i(k), k \in \mathbb{Z}_+ \tag{12}$$

Algorithm 2: A Deterministic Greedy-Based Algorithm to Construct a Sparse Weighted Actuator Schedule (Theorem 1).

Input: $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, t and d
Output: $s_i(k) \geq 0$ for $(i, k+1) \in [m] \times [t]$
1: $\mathcal{C}(t) := [B\ AB\ A^2B\ \cdots\ A^{t-1}B]$
2: Set $V = (\mathcal{C}(t)\mathcal{C}^\top(t))^{-\frac{1}{2}}\mathcal{C}(t)$
3: Set $U = V$
4: Run $[c_1, \ldots, c_{mt}] = \text{DualSet}^*(V, U, dt)$
5: **return** $s_i(k) := \sqrt{c_{i+mk}/(1+\frac{n}{dt})}$ for $(i, k+1) \in [m] \times [t]$

where $s_i(k) \geq 0$ shows the strength of the ith control input at time k. The controllability Gramian (5) at time t for this system can be rewritten as

$$\mathcal{W}_s(t) = \sum_{k=0}^{t-1}\sum_{j\in\sigma_k} s_j^2(k)\left(A^{t-k-1}b_j\right)\left(A^{t-k-1}b_j\right)^\top. \tag{13}$$

Our goal is to reduce the number of active actuators *on average* d, where

$$d := \frac{\sum_{k=0}^{t-1}\text{card}\,\{\sigma_k\}}{t} \tag{14}$$

such that the controllability Gramian of the fully actuated and the new sparsely actuated system are "close." Of course, this approximation will require horizon lengths that are potentially longer than the dimension of the state. The definition below formalizes this approximation.

Definition 2: Given a time horizon $t \geq n$, system (12) with a weighted actuator schedule is (ϵ, d)-approximation of system (1) if and only if

$$(1-\epsilon)\,\mathcal{W}(t) \preceq \mathcal{W}_s(t) \preceq (1+\epsilon)\,\mathcal{W}(t) \tag{15}$$

where $\mathcal{W}(t)$ and $\mathcal{W}_s(t)$ are the controllability Gramian matrices of (1) and (12), respectively, and parameter d is defined by (14) as the average number of active actuators, and $\epsilon \in (0, 1)$ is the approximation factor.

Remark 2: While it might appear that allowing for the choice of $s_i(k)$ might lead to amplification of input signals, we note that the scaling cannot be too large because the approximation is two-sided. Specifically, by taking the trace from both sides of (15), we can see that the weighted summation of $s_i^2(k)$'s is bounded. Moreover, based on Definition 2, the ranks of matrices $\mathcal{W}(t)$ and $\mathcal{W}_s(t)$ are the same. Thus, the resulting (ϵ, d)-approximation remains controllable (recall that we assume that the original system is controllable).

Remark 3: The results presented in this article also work for the case of linear time-varying systems, and it is straightforward to extend them for affine nonlinear discrete-time systems as well.

A. Deterministic Approach: Sparsifying Sums of Rank-One Matrices

The following theorem constructs a solution for the sparse weighted actuator schedule problem in polynomial time.

Theorem 1: Given the time horizon $t \geq n$, model (1), and $d > 1$, Algorithm 2 deterministically constructs an actuator schedule such that the resulting system (12) is a (ϵ, d)-approximation of (1) with $\epsilon = \frac{2}{\sqrt{\frac{dt}{n}}+\sqrt{\frac{n}{dt}}}$ in at most $\mathcal{O}(dm(tn)^2)$ operations.

Proof: The controllability Gramian of (1) at time t is given by

$$\mathcal{W}(t) = \sum_{i=0}^{t-1}\sum_{j=1}^{m}(\underbrace{A^i b_j}_{v_{ij}})(A^i b_j)^\top = \sum_{i=0}^{t-1}\sum_{j=1}^{m} v_{ij} v_{ij}^\top. \tag{16}$$

By multiplying $\mathcal{W}^{-\frac{1}{2}}(t)$ on both sides of (16), it follows:

$$I = \sum_{i=0}^{t-1}\sum_{j=1}^{m}\underbrace{(\mathcal{W}^{-\frac{1}{2}}(t)A^i b_j)}_{\bar{v}_{ij}}(\mathcal{W}^{-\frac{1}{2}}(t)A^i b_j)^\top = \sum_{i=0}^{t-1}\sum_{j=1}^{m}\bar{v}_{ij}\bar{v}_{ij}^\top. \tag{17}$$

Next, we define $U := \{\bar{v}_{ij} | i+1 \in [t], j \in [m]\}$ and $V := U$. According to (3), (16), and (17), elements of U are the columns of matrix $(\mathcal{C}(t)\mathcal{C}^\top(t))^{-\frac{1}{2}}\mathcal{C}(t)$. We now apply Lemma 1, which shows that there exist scalars $\bar{c}_{ij} \geq 0$ with

$$\text{card}\,\{(i,j) \,:\, i+1 \in [t], j \in [m], \bar{c}_{ij} > 0\} \leq \frac{dt}{n} \times n \tag{18}$$

such that

$$\left(1-\sqrt{\frac{n}{dt}}\right)^2 I \preceq \sum_{i=0}^{t-1}\sum_{j=1}^{m}\bar{c}_{ij}\,\bar{v}_{ij}\,\bar{v}_{ij}^\top$$

and

$$\sum_{i=0}^{t-1}\sum_{j=1}^{m}\bar{c}_{ij}\,\bar{v}_{ij}\,\bar{v}_{ij}^\top \preceq \left(1+\sqrt{\frac{n}{dt}}\right)^2 I$$

or equivalently

$$\left(1-\sqrt{\frac{n}{dt}}\right)^2 \mathcal{W}(t) \preceq \sum_{i=0}^{t-1}\sum_{j=1}^{m}\bar{c}_{ij}\,v_{ij}\,v_{ij}^\top \preceq \left(1+\sqrt{\frac{n}{dt}}\right)^2 \mathcal{W}(t). \tag{19}$$

We can of course write the controllability Gramian of (12) at time t as

$$\mathcal{W}_s(t) = \sum_{i=0}^{t-1}\sum_{j=1}^{m}s_j^2(t-i-1)(\underbrace{A^i b_j}_{v_{ij}})(A^i b_j)^\top = \sum_{i=0}^{t-1}\sum_{j=1}^{m}s_j^2(t-i-1)\,v_{ij}v_{ij}^\top.$$

Define $\epsilon := \frac{2}{\sqrt{\frac{dt}{n}}+\sqrt{\frac{n}{dt}}}$, and

$$s_j(t-i-1) := \sqrt{\bar{c}_{ij}/(1+\frac{n}{dt})}. \tag{20}$$

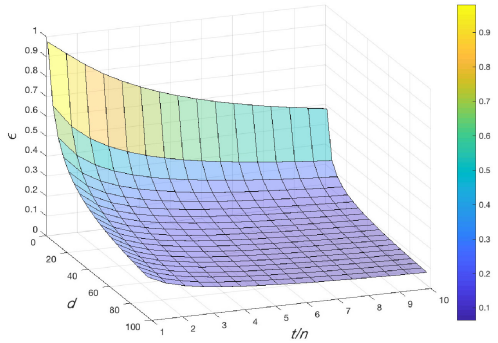


Fig. 1. This plot presents the approximation factor ϵ given by Theorem 1 versus the average number of active actuators $d \in (1, 100]$ and the normalized time horizon t/n.

Then, by substituting $(1+\frac{n}{dt})s_j^2(t-i-1)$ for $\bar{c}_{ij}$ in (19), we get

$$(1-\epsilon)\mathcal{W}(t) \preceq \mathcal{W}_s(t) \preceq (1+\epsilon)\mathcal{W}(t). \tag{21}$$

Finally, using (21), (18), and Definition 2, we obtain the desired result. Moreover, this algorithm runs in dt iterations. In each iteration, the functions $\mathfrak{U}$ and $\mathfrak{L}$ are evaluated at most mt times. All mt evaluations for both functions need at most $\mathcal{O}(n^3 + mtn^2)$ time, because for all of them the matrix inversions and eigenvalue decompositions can be calculated once. Finally, the updating step needs an additional $\mathcal{O}(n^2)$ time. Overall, the complexity of the algorithm is of the order $\mathcal{O}(dm(tn)^2)$. ■

Remark 4: For a given $d \geq 1$, while choosing dt columns of the controllability matrix that form a full row rank matrix (i.e., the system is controllable) is an easy task but finding dt columns of the controllability matrix that approximate the full Gramian matrix is what we are interested in here. To do so, we should note that approximating the full Gramian matrix while keeping the number of active actuators less than a constant d at each time is not possible in general. For example, in the case that $A = \mathbf{0}_{n\times n}$ and $B = I_n$, at least all actuators at time $k = t-1$ are needed to form a full row rank matrix (or to approximate the full Gramian matrix). However, as we mentioned earlier, the number of active actuators on average can be kept constant in order to approximate the full Gramian matrix. Furthermore, condition $dt \geq n$ is needed for any algorithm that has a hope of success. Indeed, taking $B = I_n$ and $A = I_n$, it is straightforward to see that if $dt < n$, then we cannot hope to approximate the controllability Gramian because the controllability matrix of any schedule with d active actuators on average is not full rank.

1) Tradeoffs: Theorem 1 illustrates a tradeoff between the average number of active actuators d and the time horizon t (also known as the time-to-control). This implies that the reduction in the average number of active actuators comes at the expense of increasing time horizon t in order to get the same approximation factor ϵ. Moreover, the approximation becomes more accurate as t and d are increased. Of course, increasing d will require more active actuators and larger t requires a larger control time window.

Fig. 1 depicts the approximation ratio ϵ given by Theorem 1 versus the average number of active actuators d and the

Algorithm 3: A Deterministic Greedy-Based Algorithm to Construct a Sparse Weighted Actuator Schedule (Corollary 1).

Input: $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, t and d
Output: $s_i(k) \geq 0$ for $(i, k+1) \in [m] \times [t]$
1: $\mathcal{C}(t) := [B\ AB\ A^2B\ \cdots\ A^{t-1}B]$
2: Set $V = (\mathcal{C}(t)\mathcal{C}^\top(t))^{-\frac{1}{2}}\mathcal{C}(t)$
3: Set

$$U = \left[\underbrace{\mathrm{e}_1, \ldots, \mathrm{e}_{mt}}_{=I_{mt}}\right]$$

4: // where $\mathrm{e}_i \in \mathbb{R}^{mt}$ for $i \in [mt]$ are the standard basis vectors for $\mathbb{R}^{mt}$
5: Run $[c_1, \ldots, c_{mt}] = \mathrm{DualSet}^*(V, U, dt)$
6: **return** $s_i(k) := \sqrt{c_{i+mk}}$ for $(i, k+1) \in [m] \times [t]$

Algorithm 4: A Deterministic Greedy-Based Algorithm to Construct a Sparse Weighted Actuator Schedule (Corollary 2).

Input: $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, t and d
Output: $s_i(k) \geq 0$ for $(i, k+1) \in [m] \times [t]$
1: $\mathcal{C}(t) := [B\ AB\ A^2B\ \cdots\ A^{t-1}B]$
2: Set $V = (\mathcal{C}(t)\mathcal{C}^\top(t))^{-\frac{1}{2}}\mathcal{C}(t)$
3: Set

$$U = \frac{1}{\sqrt{t}}\left[\underbrace{\mathrm{e}_1, \ldots, \mathrm{e}_m}_{=I_m}, \ldots, \underbrace{\mathrm{e}_1, \ldots, \mathrm{e}_m}_{=I_m}\right]$$

4: // where $\mathrm{e}_i \in \mathbb{R}^{m}$ for $i \in [m]$ are the standard basis vectors for $\mathbb{R}^{m}$ and $UU^\top = I_m$
5: Run $[c_1, \ldots, c_{mt}] = \mathrm{DualSet}^*(V, U, dt)$
6: **return** $s_i(k) := \sqrt{c_{i+mk}}$ for $(i, k+1) \in [m] \times [t]$

normalized time horizon t/n. We note that the approximation factor improves as t become larger than n. Moreover, because of $\frac{2}{x+\frac{1}{x}} \leq 1$ for $x > 0$, the approximation factor $\epsilon = \frac{2}{\sqrt{\frac{dt}{n}}+\sqrt{\frac{n}{dt}}}$ is always less than or equal to one. Hence, the upper bound ratio in (15) is at most two.

2) Sparse Actuator Schedules With Energy Constraints: In this section, based on the energy/budget constraints on the scalings $s_i(k)$'s where $i \in [m]$ and $k+1 \in [t]$; three cases are considered as follows.

1) The scaling ratios are bounded, i.e.,

$$\max_{i\in[m],k+1\in[t]} s_i^2(k) \leq \gamma.$$

2) The sum of scaling ratios for each input is bounded, i.e.,

$$\max_{i\in[m]} \sum_{k+1\in[t]} s_i^2(k) \leq \gamma.$$

3) The sum of scaling ratios at each time is bounded, i.e.,

$$\max_{k+1\in[t]} \sum_{i\in[m]} s_i^2(k) \leq \gamma.$$

In the following corollaries, we present deterministic sparse actuator schedules with the abovementioned energy/budget constraints. These corollaries trade one of the inequalities in Theorem 1 with a single fixed bound on the size of scalings.

Corollary 1: Given the time horizon $t \geq n$, model (1), and $d > 1$, Algorithm 3 deterministically constructs an actuator schedule for (12) in at most $\mathcal{O}(dm(tn)^2)$ operations such that, on average, at most d active actuators are selected, and the following bound:

$$\rho(\mathcal{W}_s(t)) \leq \left(1 - \sqrt{\frac{n}{dt}}\right)^{-2} \rho(\mathcal{W}(t))$$

holds for all systemic controllability measures. Moreover, the maximum scaling ratio over all time and inputs is bounded by

$$\max_{i\in[m],k+1\in[t]} s_i^2(k) \leq \gamma$$

where $\gamma = (1 + \sqrt{\frac{m}{d}})^2$.

Proof: The proof is a simple variation on the proof of Theorem 1, and is not repeated here. ■

As expected, the abovementioned result shows that the scaling becomes smaller as the ratio m/d decreases.

Corollary 2: Given the time horizon $t \geq n$, model (1), and $d > 1$, Algorithm 4 deterministically constructs an actuator schedule for (12) in $\mathcal{O}(dm(tn)^2)$ operations such that it has, on average, at most d active actuators, and the following:

$$\rho(\mathcal{W}_s(t)) \leq \left(1 - \sqrt{\frac{n}{dt}}\right)^{-2} \rho(\mathcal{W}(t))$$

holds for all systemic controllability measures. Moreover, the sum of scaling ratios for all inputs is bounded by

$$\max_{i\in[m]} \sum_{k=0}^{t-1} s_i^2(k) \leq \gamma$$

where $\gamma = t(1 + \sqrt{\frac{m}{dt}})^2$.

Proof: The proof is a simple variation on the proof of Theorem 1, and is not repeated here. ■

Corollary 3: Given the time horizon $t \geq n$, model (1), and $d > 1$, Algorithm 5 deterministically constructs an actuator schedule for (12) in $\mathcal{O}(dm(tn)^2)$ operations such that it has, on average, at most d active actuators, and the following:

$$\rho(\mathcal{W}_s(t)) \leq \left(1 - \sqrt{\frac{n}{dt}}\right)^{-2} \rho(\mathcal{W}(t))$$

holds for all systemic controllability measures. Moreover, the sum of scaling ratios at each time is bounded by

$$\max_{k+1\in[t]} \sum_{i=1}^{m} s_i^2(k) \leq \gamma$$

where $\gamma = m(1 + \sqrt{\frac{1}{d}})^2$.

Proof: The proof is a simple variation on the proof of Theorem 1, and is not repeated here. ■

Algorithm 5: A Deterministic Greedy-Based Algorithm to Construct a Sparse Weighted Actuator Schedule (Corollary 3).

Input: $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, t and d
Output: $s_i(k) \geq 0$ for $(i, k+1) \in [m] \times [t]$

1: $\mathcal{C}(t) := [B\ AB\ A^2B\ \cdots\ A^{t-1}B]$
2: Set $V = (\mathcal{C}(t)\mathcal{C}^\top(t))^{-\frac{1}{2}}\mathcal{C}(t)$
3: Set

$$U = \frac{1}{\sqrt{m}}\left[\underbrace{\mathrm{e}_1,\ldots,\mathrm{e}_1}_{m\text{times}},\ldots,\underbrace{\mathrm{e}_t,\ldots,\mathrm{e}_t}_{m\text{times}}\right]$$

4: // where $\mathrm{e}_i \in \mathbb{R}^t$ for $i \in [t]$ are the standard basis vectors for $\mathbb{R}^t$ and $UU^\top = I_t$
5: Run $[c_1,\ldots,c_{mt}] = \mathrm{DualSet}^*(V, U, dt)$
6: **return** $s_i(k) := \sqrt{c_{i+mk}}$ for $(i, k+1) \in [m] \times [t]$

We use a different idea in Section V-B, to develop scalable algorithms that sparsify control inputs by employing a subsampling method for a time-varying actuator schedule. This, however, come at the cost of an extra log factor in terms of the average number of selected actuators.

B. Randomized Approach: Sampling Based on the Leverage Score

In this section, we focus on a computationally tractable method for the weighted sparse actuator scheduling problem that achieve near optimal solution.

Definition 3: The leverage score of the ith column of matrix $P \in \mathbb{R}^{n\times m}$ is defined as

$$\ell_i = p_i^\top (PP^\top)^\dagger p_i$$

where p_i is the ith column of matrix P.

This quantity encodes the importance of the ith column compared to the other columns. A larger leverage score shows that the corresponding column has more influence on the spectrum of P. Based on the leverage score definition, we get $\ell_i \in [0, 1]$ for all $i \in [m]$. Because ℓ_i's are the diagonal elements of the projection matrix $P^\top(PP^\top)^{-1}P$ and the diagonal elements of the projection matrix are between zero and one. Leverage score $\ell_i = 1$ means that the ith column has a component orthogonal to the rest of the columns. Therefore, eliminating that column will decrease the rank of matrix P. On the other hand, $\ell_i = 0$ means that the ith column is parallel to the rest of the columns. When the corresponding matrix is the graph Laplacian, this quantity reduces to the effective resistance of each link in a graph [26].

We group the columns of $\mathcal{C}(t)$ in the following form:

$$\mathcal{C}(t) = \left[\underbrace{[b_1 Ab_1 \cdots A^{t-1}b_1]}_{\mathcal{C}_1(t)} \cdots \underbrace{[b_m Ab_m \cdots A^{t-1}b_m]}_{\mathcal{C}_m(t)}\right]$$

where b_j is the jth column of matrix B. Matrix $\mathcal{C}_j(t)$ presents the controllability matrix of input j at time t. The leverage score

Algorithm 6: A simple randomized algorithm to compute a sparse weighted actuator schedule $\{\sigma_i\}_{i=0}^{t-1}$ (Theorem 2).

Input : $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, t and d
Output: $\{\sigma_i\}_{i=0}^{t-1}$ and $s_i(k-1)$ for $(i, k) \in [m] \times [t]$

$\mathcal{C}(t) := \left[B\ AB\ A^2B\ \cdots\ A^{t-1}B\right]$
set $\{\sigma_i\}_{i=0}^{t-1}$ to be the empty sets (*i.e.* $\sigma_i := \{\}$)
set $s_i(k-1) = 0$ for $(i, k) \in [m] \times [t]$
set $\pi(i,k) = \frac{\mathrm{trace}\left((\mathcal{C}(t)\mathcal{C}^\top(t))^\dagger A^{t-k}b_i(A^{t-k}b_i)^\top\right)}{n}$ for all $(i, k) \in [m] \times [t]$
for $j = 1$ *to* $M := \lceil dt \rceil$ **do**
  $(i, k) \leftarrow$ sample (i, k) from $[m] \times [t]$ with probability distribution π
  $\sigma_{k-1} = \sigma_{k-1} \cup \{i\}$
  $s_i^2(k-1) = s_i^2(k-1) + \frac{1}{M\pi(i,k)}$
end
return $\{\sigma_i\}_{i=0}^{t-1}$ and $s_i(k-1)$ for $(i, k) \in [m] \times [t]$

for each column of $\mathcal{C}(t)$ is defined as

$$\ell(A^i b_j) = (A^i b_j)^\top \left(\mathcal{C}(t)\,\mathcal{C}^\top(t)\right)^\dagger A^i b_j \tag{22}$$

where $(i+1) \in [t]$ and $j \in [m]$. For these scores, we have

$$\begin{aligned}\sum_{i=0}^{t-1}\sum_{j=1}^{m}\ell(A^i b_j) &= \mathrm{trace}\left(\mathcal{C}^\top(t)(\mathcal{C}(t)\mathcal{C}^\top(t))^\dagger\mathcal{C}(t)\right)\\ &= \mathrm{trace}\left((\mathcal{C}(t)\mathcal{C}^\top(t))^\dagger\mathcal{C}(t)\mathcal{C}^\top(t)\right)\\ &= \mathrm{trace}(I_n) = n.\end{aligned} \tag{23}$$

In (23), we use the fact that $\mathrm{trace}(AB) = \mathrm{trace}(BA)$ (i.e., the matrices in a trace of a product can be switched without changing the result as long as A and $B^\top$ have the same dimensions), and $\mathrm{rank}(\mathcal{C}(t)) = n$ (i.e., the system is controllable).

We now randomly sample the actuators with probabilities proportional to their leverage scores to sparsify control inputs. This sampling occurs across time and over all possible actuators at each time (see Algorithm 6). At every time, each actuator is kept active or inactive according to probability $\ell(A^i b_j)/n$, where $(i+1) \in [t]$ and $j \in [m]$. Using [26, Th. 1], we can construct a sampling strategy that utilizes the leverage score to probabilistically choose actuators. The catch is that there is an extra $\log n$ factor in the average number of selected actuators, and potentially different actuators are chosen at different times.

Theorem 2: Assume that dynamics (1), time horizon $t \geq n$, and approximation factor $\epsilon \in [1/\sqrt{n}, 1)$ are given. Choose a real number $d = \frac{9c^2 n\log n}{t\epsilon^2}$, where c is the constant in Lemma 2. Then, Algorithm 6 produces scheduling (12), which is (ϵ, d)-approximation of (1) with probability of at least 0.5 for sufficiently large n.[5]

[5]We should note that one can repeat Algorithm 6 for example $r = 4$ times to get the desired results with more than $1 - \frac{1}{2^r} = 0.9375$ probability. Moreover, the probability is improved by increasing the number of iteration r. Assume r is a constant and does not depend on n, t, d, or m. Then, repeating the algorithm

Proof: The structure of the proof follows from the proof of [26, Th. 4]. Let us start with the following projection matrix:

$$\Pi = \mathcal{C}(t)^\top \mathcal{W}^{-1}(t)\, \mathcal{C}(t) \tag{24}$$

where $\mathcal{C}(t)$ is n-by-tm controllability matrix (2) and matrix $\mathcal{W}(t) = \mathcal{C}(t)\, \mathcal{C}^\top(t)$ is given by (3). The tm-by-tm projection matrix Π has eigenvalue at 0 with multiplicity $t \times m - n$ and eigenvalue at 1 with multiplicity n. Therefore, we get

$$\operatorname{trace}(\Pi) = \operatorname{rank}(\Pi) = n. \tag{25}$$

The set X is obtained based on columns of Π as follows:

$$X = \left\{ y_j \in \mathbb{R}^n : y_j = (\pi_j)^{-1/2}\, \Pi(.,j), \text{and} j \in [tm] \right\}$$

where matrix Π is given by (24), vector $\Pi(.,j)$ is the jth column of Π, and π_j is probability of selecting vector y_j (i.e., $\pi(y_j) = \pi_j$). The probability distribution π over X is defined by

$$\pi(y_j) := \pi_j = \frac{\Pi(j,j)}{\operatorname{rank}(\Pi)} = \frac{\Pi(j,j)}{n} \tag{26}$$

where $\Pi(j,j)$ is the jth diagonal element of matrix Π and $j \in [tm]$. Based on (24), each columns of Π corresponds to the ith input at time $k-1$ where $(i,k) \in [m] \times [t]$. The mapping comes from the controllability matrix structure (2), and we define it as $\mathfrak{m}(.) : [mt] \to [m] \times [t]$, where

$$\mathfrak{m}(j) = (i,k) = \left(j - t\lfloor \frac{j}{t} \rfloor, t - \lfloor \frac{j}{m} \rfloor \right). \tag{27}$$

This means the jth column of Π corresponds to (i,k), where $k = t - \lfloor \frac{j}{m} \rfloor$ and $i = j - t\lfloor \frac{j}{t} \rfloor$. Thus, in Algorithm 6, for notational simplicity we denote $\pi(y_j) := \pi(\mathfrak{m}(j)) = \pi(i,k)$. For each element of X, we have

$$\begin{aligned} \|y_j\| &= (\pi_j)^{-1/2}\, \|\Pi(.,j)\| = \left(\frac{n}{\Pi(j,j)} \right)^{1/2} \times (\Pi(j,j))^{1/2} \\ &= \sqrt{n} \end{aligned} \tag{28}$$

where we use the fact that

$$\Pi(.,j)^\top \Pi(.,j) = \Pi(j,j)$$

because Π is an orthogonal projection matrix (i.e., $\Pi\Pi = \Pi$). Then, using (25) and (28), we have

$$\begin{aligned} \mathbf{E}(yy^\top) &= \sum_{j=1}^{tm} \pi_j y_j y_j^\top = \sum_{j=1}^{tm} \pi_j \frac{1}{\pi_j} \Pi(.,j)\Pi(.,j)^\top \\ &= \Pi\Pi = \Pi \end{aligned} \tag{29}$$

where y is a random variable vector with a countable set of outcomes X occurring with probabilities π defined by (26). Let $\hat{y}_1, \ldots, \hat{y}_M$ be independent samples drawn from π, then, based on Algorithm 6, then we have

$$\begin{aligned} \Pi\Gamma\Pi &= \sum_{j=1}^{tm} \Gamma(j,j)\Pi(.,j)\Pi(.,j)^\top \\ &= \sum_{j=1}^{tm} s_i^2(k-1)\Pi(.,j)\Pi(.,j)^\top \\ &= \sum_{j=1}^{tm} \frac{\#\text{ of times } (i,k) \text{ is sampled}}{M\pi(i,k)} \Pi(.,j)\Pi(.,j)^\top \\ &= \frac{1}{M} \sum_{j=1}^{tm} \frac{\#\text{ of times } (i,k) \text{ is sampled}}{\pi(i,k)} \Pi(.,j)\Pi(.,j)^\top \\ &= \frac{1}{M} \sum_{j=1}^{M} \hat{y}_j \hat{y}_j^\top \end{aligned} \tag{30}$$

where Γ is a nonnegative diagonal matrix and the random entry $\Gamma(j,j)$ specifies the "amount" of the ith input at time $k-1$ (where $\mathfrak{m}(j) = (i,k)$) included in the sparse actuator scheduling by Algorithm 6. For instance, $\Gamma(j,j) = 1/M\pi(\mathfrak{m}(j))$ if he ith input at time $k-1$ is sampled once, $2/M\pi(\mathfrak{m}(j))$ if it is sampled twice, and zero if it is not sampled at all. The scaling of the ith input at time $k-1$ in the scheduling is given by $s_i^2(k-1) = \Gamma(j,j)$, where $\mathfrak{m}(j) = (i,k)$. We next use a concentration lemma to prove this theorem. Using Lemma 2, (28), (29), and (30), we get

$$\mathbf{E} \left\| \frac{1}{M} \sum_{i=1} \hat{y}_i \hat{y}_i^\top - \mathbf{E}\, yy^\top \right\|_2 = \mathbf{E}\, \|\Pi\Gamma\Pi - \Pi\|_2 \tag{31}$$

$$\leq \min\left(1, c\sqrt{\frac{n \log M}{M}}\right) \tag{32}$$

where c is an absolute constant. Assuming $M = 9c^2 n \log n/\epsilon^2$ gives[6]

$$\begin{aligned} \mathbf{E}\, \|\Pi\Gamma\Pi - \Pi\|_2 &\leq c\sqrt{\frac{n \log M}{M}} \\ &\leq \epsilon \sqrt{\frac{\log(9c^2 n \log n/\epsilon^2)}{9 \log n}} \leq \epsilon/2 \end{aligned} \tag{33}$$

for n sufficiently large, and ϵ is assumed to be in $[1/\sqrt{n}, 1)$. By Markov's inequality and (33), we have

$$\mathbf{Pr}\left[\|\Pi\Gamma\Pi - \Pi\| > \epsilon \right] \leq 0.5$$

which means we have

$$\|\Pi - \Pi\Gamma\Pi\|_2 \leq \epsilon \tag{34}$$

with probability of at least 0.5. Note that Γ is a nonnegative diagonal matrix with weights $s_i^2(k)$ on its diagonal such that $\mathcal{W}_s(t) = \mathcal{C}(t)\, \Gamma\, \mathcal{C}^\top(t)$. Based on [26, Lemma 4], the inequality (34) is equivalent to

$$\sup_{\substack{x \in \mathbb{R}^{tm} \\ x \neq 0}} \frac{|x^\top(\Pi - \Pi\Gamma\Pi)x|}{x^\top x} \leq \epsilon. \tag{35}$$

r times does not change the time complexity of the algorithm, and we still have an approximately linear time algorithm. Therefore, by repeating the algorithm and choosing the best result, we can obtain the same error bound with higher probability.

[6]It can be shown that $M = 4n \log n\epsilon^2$ would be enough to get ϵ approximation with high probability [46].

Since we have $\mathrm{Im}\{\mathcal{C}^\top(t)\} \subset \mathbb{R}^{mt}$, it follows:

$$\sup_{\substack{x\in \mathrm{Im}\{\mathcal{C}^\top(t)\}\\ x\neq 0}} \frac{|x^\top(\Pi - \Pi\Gamma\Pi)x|}{x^\top x} \leq \sup_{\substack{x\in \mathbb{R}^{mt}\\ x\neq 0}} \frac{|x^\top(\Pi - \Pi\Gamma\Pi)x|}{x^\top x} \leq \epsilon.$$

Let us define $x = \mathcal{C}^\top(t)x'$. Then, we rewrite (35) as follows:

$$\sup_{\substack{x'\in\mathbb{R}^n\\ x'\notin \ker\{\mathcal{C}^\top(t)\}}} \frac{|x'^\top(\mathcal{W}(t) - \mathcal{W}_s(t))x'|}{x'^\top \mathcal{W}(t)x'} \leq \epsilon. \tag{36}$$

As a result, it follows:

$$\sup_{\substack{x'\in\mathbb{R}^n\\ x'\neq 0}} \frac{|x'^\top(\mathcal{W}(t) - \mathcal{W}_s(t))x'|}{x'^\top \mathcal{W}(t)x'} \leq \epsilon \tag{37}$$

which implies that

$$(1-\epsilon)\mathcal{W}(t) \preceq \mathcal{W}_s(t) = \mathcal{C}(t)\Gamma\mathcal{C}^\top(t) \preceq (1+\epsilon)\mathcal{W}(t). \tag{38}$$

Finally, using (38) and Definition 2, it is straightforward to show that for every systemic controllability measure $\rho : \mathbb{S}^n_+ \to \mathbb{R}_+$, we have

$$\left|\frac{\rho(\mathcal{W}(t)) - \rho(\mathcal{W}_s(t))}{\rho(\mathcal{W}(t))}\right| \leq \epsilon.$$

Therefore, we conclude the desired result. ■

This result shows that with a simple randomized sampling strategy, one can choose on average less than $\mathcal{O}(\log n/\epsilon^2)$ number of actuators at each time, to approximate any of the controllability metrics when $t = n$. Moreover, this result shows that it is possible to have a time-varying actuator schedule with a constant number of active actuators on average over a time horizon a little longer than n (i.e., $t = \mathcal{O}(n \log n)$) via random sampling. Algorithm 6 computes the sparse actuator schedule using a nearly-linear time $\tilde{\mathcal{O}}(mt)$ algorithm[7] with guaranteed performance bounds, where mt is the total number of actuations (time-to-control × number of inputs). This favorable almost-linear-time complexity is achieved by random sampling of actuators in both time and domain based on their leverage scores [26]. According to Theorem 1, the average number of active actuators can be reduced to $\mathcal{O}(1/\epsilon^2)$, at the expense of either solving SDPs [25] or greedily handling certain eigenvalue bounds (see Algorithm 2). Algorithm 6 is conceptually simpler than Algorithm 2 and the SDP-based algorithm presented in [25], which provide $d = \mathcal{O}(1/\epsilon^2)$ in $\mathcal{O}(m(tn)^2/\epsilon^2)$ and $\tilde{\mathcal{O}}(mt/\epsilon^{\mathcal{O}(1)})$ time, respectively.

The concept of a leverage score for each column can be generalized to a group of columns as follows:

$$\ell_{\mathcal{C}_i} = \mathrm{trace}\left(\mathcal{C}_i^\top(t)\left(\mathcal{C}(t)\,\mathcal{C}^\top(t)\right)^\dagger \mathcal{C}_i(t)\right). \tag{39}$$

Using group leverage scores, one can also use a greedy heuristic algorithm to obtain an approximation solution for the static scheduling problem. We note that the problem of approximation of the controllability Gramian with a sparse, static actuator set is considerably more challenging as it does not lend itself to a sampling-based strategy: any choice made at one time has to be consistent with the next.

When using a time-varying schedule, the contribution of each actuator to the Gramian at each time is a rank-one matrix. Therefore, we can use the machinery developed for the Kadison–Singer conjecture to find a sparse subset of actuators over time to approximate the (potentially very large) sum of rank-one matrices. In the static case, however, the choices of actuators at different times are all the same. As a result, the Gramian can be written as a sum of positive semidefinite matrices corresponding to the selected actuators at each time. Finding a sparse approximation in this case would require a generalization of the Kadison–Singer conjecture from sums of rank-one to sums of higher ranked positive semidefinite matrices. Such a result has remained elusive as of yet.

VI. Unweighted Sparse Actuator Schedule

In the previous section, we allowed for rescaling of the input to come up with a sparse approximation of the Gramian. Here, we assume that the actuator/signal strength cannot be arbitrarily set for individual active actuators and only can be 0 or 1. Given a time horizon $t \geq n$, our problem is to compute an actuator schedule $\sigma = \{\sigma_k\}_{k=0}^{t-1}$, where $\sigma_k \subset [m]$ for the system (1), i.e.,

$$x(k+1) = A\,x(k) + \sum_{i\in\sigma_k} b_i\,u_i(k), k \in \mathbb{Z}_+. \tag{40}$$

As before, the controllability Gramian at time t for schedule (40) is given by

$$\mathcal{W}_\sigma(t) := \sum_{i=0}^{t-1}\sum_{j\in\sigma_i}(A^{t-i-1}b_j)(A^{t-i-1}b_j)^\top. \tag{41}$$

Optimal actuator selection can, now, be formulated as a combinatorial optimization problem. We consider both static and dynamic actuator schedules, corresponding to time-invariant and time-varying input matrices.

1) Static Scheduling Problem: In this case, all sets $\sigma_i \subset [m]$ for $i+1 \in [t]$ are identical, which means we keep the same schedule at every point in time for the whole time horizon t

$$\min_{\sigma\in\mathcal{S}(m,d_{\max})} \rho\left(\sum_{i=0}^{n-1}\sum_{j\in\sigma}(A^ib_j)(A^ib_j)^\top\right) \tag{42}$$

where

$$\mathcal{S}(m,d_{\max}) := \{\sigma : \sigma \subset [m], \mathrm{card}(\sigma) \leq d_{\max}\} \tag{43}$$

where $d_{\max}$ is a given upper bound on the number of active actuators at each time, and m is the total number of actuators.

2) Time-Varying Scheduling Problem: In this case, the optimal dynamic strategy is given as

$$\min_{\{\sigma_i\}_{i=0}^{t-1}\in\mathcal{S}(m,d_{\max},t)} \rho\left(\sum_{i=0}^{t-1}\sum_{j\in\sigma_i}(A^{t-i-1}b_j)(A^{t-i-1}b_j)^\top\right) \tag{44}$$

[7] $f(n) \in \tilde{\mathcal{O}}(g(n))$ means that there exists $c > 0$ such that $f(n) \in \mathcal{O}(g(n)\log^c g(n))$.

where

$$\mathcal{S}(m, d_{\max}, t) := \left\{ \{\sigma_i\}_{i=0}^{t-1} : \sigma_i \subset [m], \sum_{i=0}^{t-1} \operatorname{card}(\sigma_i) \leq t d_{\max} \right\} \tag{45}$$

and $d_{\max}$ is a given upper bound on the average number of active actuators at each time, i.e., $d_{\max} \geq \sum_{i=0}^{t-1} \operatorname{card}(\sigma_i)/t$, where t is a time horizon, and m is the total number of actuators.

The exact combinatorial optimization problems (42) and (44) are intractable and NP-hard optimization problems; however, it is straightforward to solve a continuous relaxation of these optimization problems where the cost function ρ is convex. To find a near-optimal solution of optimization problems (42) and (44), one can use a variety of standard methods for optimal experimental design (greedy methods, sampling methods, the classical pipage rounding method combined with SDP). Specifically, in the case of submodular systemic controllability measures (e.g., D- and T-optimality), the classical rounding method (e.g., pipage and randomized rounding) combined with SDP relaxation results in computationally fast algorithms with a constant approximation ratio [44]. These approaches are not applicable to nonsubmodular systemic measures, such as A-, and E-optimality [24], [47].

In the following result, we use a result based on regret minimization of the least eigenvalues of positive semidefinite matrices (cf. [24]) to obtain a constant approximation ratio for all systemic controllability metrics.

Theorem 3: Assume that time horizon $t \geq n$, dynamics (1), systemic controllability metric $\rho : \mathbb{S}_+^n \to \mathbb{R}$, and $d_{\max} > 2$ are given. Then, there exists a polynomial-time algorithm, which computes a schedule $\hat{\sigma} = \{\hat{\sigma}_i\}_{i=0}^{t-1}$ that satisfies

$$\rho(\mathcal{W}_{\hat{\sigma}}(t)) \leq \gamma\left(\frac{d_{\max} t}{n}\right) . \min_{\{\sigma_i\}_{i=0}^{t-1} \in \mathcal{S}(m, d_{\max}, t)} \rho\left(\mathcal{W}_{\sigma}(t)\right)$$

where $\gamma(d_{\max} t/n)$ is a positive constant depending only on $d_{\max} t/n$.

Proof: The proof is a simple variation on the proof of [24, Th. 1.1], and is not repeated here. ■

The positive constant $\gamma(.)$ in Theorem 3 is defined as follows:

$$\gamma(\zeta) = \min_{y > \frac{3\zeta}{\zeta - 2}} \frac{v(2 + \frac{v}{\zeta})}{(1 - \frac{2}{\zeta})v - 3}, \quad \text{where } \zeta > 2 \tag{46}$$

see the proof of [24, Th. 1.1]. For example, for $d_{\max} t/n \in \{4, 10, 50\}$, using (46), we get

$$\gamma(4) = 2\left(5 + \sqrt{21}\right) \approx 19.1652$$

$$\gamma(10) = 5\left(\frac{11 + \sqrt{57}}{16}\right) \approx 5.79682$$

and

$$\gamma(50) = 25\left(\frac{17 + \sqrt{33}}{192}\right) \approx 2.96153.$$

Next, we use the results from Section V to obtain an unweighted sparse actuator schedule with guaranteed performance bound.

Algorithm 7: A Deterministic Greedy-Based Algorithm to Construct a Sparse Unweighted Actuator Schedule (Corollary 4).

Input: $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, t and $d_{\max}$
Output: $s_i(k)$ for $(i, k+1) \in [m] \times [t]$

1: $\mathcal{C}(t) := [B\ AB\ A^2B\ \cdots\ A^{t-1}B]$
2: Set $V = (\mathcal{C}(t)\mathcal{C}^\top(t))^{-\frac{1}{2}}\mathcal{C}(t)$
3: Set

$$U = \left[\underbrace{\mathrm{e}_1, \ldots, \mathrm{e}_{mt}}_{=I_{mt}}\right]$$

4: //where $\mathrm{e}_i \in \mathbb{R}^{mt}$ for $i \in [mt]$ are the standard basis vectors for $\mathbb{R}^{mt}$
5: Run $[c_1, \ldots, c_{mt}] = \mathrm{DualSet}^*(V, U, d_{\max} t)$
6: **return** $s_i(k) := \left\lceil \sqrt{c_{i+mk}}/(1 + \sqrt{\frac{m}{d_{\max}}}) \right\rceil$ for $(i, k+1) \in [m] \times [t]$

Algorithm 8: A greedy heuristic for given $\rho(.)$ which sequentially picks inputs $\mathrm{GreedyStatic}(A, B, t, d)$.

Input : $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, t and d
Output: $B_s \in \mathbb{R}^{n \times d}$, $\rho(\mathcal{W}_s)$

$\mathcal{W}_s := \mathbf{0}_{n \times n}$
for $k = 1$ *to* d **do**
  $j \leftarrow$ find a column of B that returns the maximum value for

$$\rho(\mathcal{W}_s + \alpha I_n) - \rho\left(\mathcal{W}_s + \sum_{i=0}^{t-1} A^i B(:, j) B(:, j)^\top (A^i)^\top + \alpha I_n\right)$$

  // $\alpha > 0$ is sufficiently small to avoid singularity
  $B_s \leftarrow [B_s, B(:, j)]$
  $\mathcal{W}_s = \sum_{i=0}^{t-1} A^i B_s B_s^\top (A^i)^\top$
  $B(:, j) \leftarrow [\,]$
end
return B_s, $\rho(\mathcal{W}_s)$

Corollary 4: Assume that time horizon $t \geq n$, dynamics (1), and $d_{\max} > 1$ are given. Then, polynomial-time Algorithm 7 deterministically constructs an actuator schedule for (12) with $s_i(k) \in \{0, 1\}$ such that it has, on average, at most $d_{\max}$ active actuators, and the following:

$$\rho(\mathcal{W}_{\sigma}(t)) \leq \left(\frac{1 + \sqrt{\frac{m}{d_{\max}}}}{1 - \sqrt{\frac{n}{d_{\max} t}}}\right)^2 \rho(\mathcal{W}(t))$$

holds for all systemic controllability measures.

Proof: The proof is a simple variation on the proof of Theorem 1, and is not repeated here. ■

In view of this result, one can choose any constant number greater than one as the number of active actuators on average to construct a sparse unweighted actuator schedule in order to approximate controllability measures. This, however, comes at the cost of an extra $(1 + \sqrt{\frac{m}{d_{\max}}})^2$ factor in terms of the energy

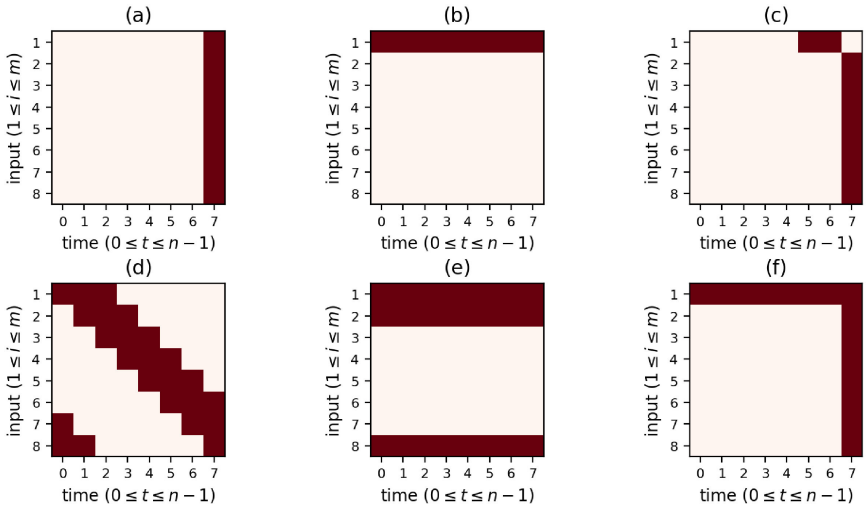


Fig. 2. Six unweighted actuator schedules for Example 1. (a) All actuators are active at time 7. (b) Actuator one is active at each time. (c) Schedule is obtained Algorithm 7. (d) Three actuators are active at all time and each actuator is used three times. (e) Three fixed actuators $\{1, 2, 8\}$ are active at all time. (f) Proposed sparse schedule based on Algorithm 7 with less than two active actuators at each time on average. The color of element (i, k) is red when $s_i(k) = 1$ and white otherwise where $i \in [8]$, $k + 1 \in [8]$ and $s_i(k) \in \{0, 1\}$. For Fig. 2(c) and (f), which are obtained based on Algorithm 7, we can observe that the actuator schedule has procrastination in actuator activations (i.e., more active actuators at the end of the time horizon); however, in Example 3 we can see "front-loaded" behavior (i.e., more active actuators early in the time horizon) due to different dynamics in this example.

Algorithm 9: A greedy heuristic for given $\rho(.)$ which sequentially picks inputs and activation times $\mathrm{GreedyTimeVarying}(A, B, t, d)$.

Input : $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, t and d

Output: $\rho(\mathcal{W}_s)$

$\mathcal{C} := \begin{bmatrix} B & AB & A^2B & \cdots & A^{t-1}B \end{bmatrix}$

$\mathcal{C}_s := \mathbf{0}_{n\times mt}$

for $k = 1$ *to* $M := \lceil dt \rceil$ **do**

  $j \leftarrow$ find a column of C that returns the maximum value for

$$\rho(\mathcal{W}_s + \alpha I_n) - \rho\left(\mathcal{W}_s + C(:, j)C(:, j)^\top + \alpha I_n\right)$$

  `// α > 0 is sufficiently small to avoid singularity`

  $C_s \leftarrow [C_s, C(:, j)]$

  $\mathcal{W}_s = C_s C_s^\top$

  $C(:, j) \leftarrow [\,]$

end

return $\rho(\mathcal{W}_s)$

cost compared to the weighted sparse actuator schedule (cf. Corollary 1).

VII. Numerical Examples

In this section, we consider three numerical examples to demonstrate the results.

We compare our results with a greedy heuristic that sequentially picks control inputs to maximize the systemic metric decrease of the controllability matrix (see Algorithm 8). The selected inputs are active at all times. It is shown that the greedy method works well and matches the inapproximability barrier[8] in polynomial time [1]. We also compare our results with a greedy algorithm for a time-varying actuator schedule that sequentially picks both control inputs and activation times to maximize the decrease in the systemic metric of the controllability Gramian (see Algorithm 9). Without loss of generality, we assume time horizon $t = n$.

Example 1 ([1]): Assume that the state-space matrices of system (1) are given by

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-7}{2} \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & \frac{-5}{2} \\ \frac{3}{4} & \frac{1}{2} & 0 & 4 & 0 & 0 & 0 & \frac{13}{8} \\ 0 & \frac{3}{4} & \frac{1}{2} & 0 & 5 & 0 & 0 & \frac{11}{8} \\ \frac{5}{4} & 0 & \frac{3}{4} & 0 & 0 & 6 & 0 & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{4} & 1 & 0 & 0 & 0 & 7 & \frac{9}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix} \tag{47}$$

and

$$B_{\min} = \mathrm{diag}\,[1, 1, 0, 0, 0, 0, 0, 1]\,. \tag{48}$$

Direct computation shows that choosing (48) makes the system controllable and no diagonal-matrix sparser than $B_{\min}$ renders A controllable. For this case ($B = B_{\min}$), the performance is

$$\mathrm{trace}\left(\sum_{i=0}^{n-1} A^i B_{\min} B_{\min}^\top (A^i)^\top\right)^{-1} = 0.503$$

[8]It approximates the minimum number of inputs in the system that need to be affected for controllability within a factor of $c \log n$ for some $c > 0$.

TABLE II
VALUES OF CONTROLLABILITY PERFORMANCE AND AVERAGE NUMBER OF ACTIVE ACTUATORS AT EACH TIME FOR THE UNWEIGHTED ACTUATOR SCHEDULE PRESENTED IN FIG. 2 AND BASED ON GREEDY ALGORITHMS 8 AND 9

	Figs. 2.(a)&(b)	Fig. 2.(c)	Fig. 2(d)	Fig. 2.(e)	Fig. 2.(f)	Algorithm 8	Algorithm 9	Fully Actuated
$\text{trace}\left(\mathcal{W}^{-1}(n)\right)$	uncontrollable	0.628	uncontrollable	0.503	0.161	uncontrollable	0.294	0.132
d	1	1.125	3	3	1.875	3	3	8

The unweighted schedules presented in Fig. 2(c) and (f) are obtained based on Algorithm 7. It is not possible to greedily select three inputs (active at all time) to make the system in Example 1 controllable.

TABLE III
VALUES OF CONTROLLABILITY PERFORMANCE FOR THREE DIFFERENT ACTUATOR SCHEDULES IN EXAMPLE 2. (1) WEIGHTED ACTUATOR SCHEDULE IN FIG. 4 BASED ON ALGORITHM 6 (2) STATIC LEADER SCHEDULE WITH 160 LEADERS ACTIVE AT ALL TIME. (3) FULLY ACTUATED CASE

	Fig. 4 (Algorithm 6)	Static Leader Schedule	Fully Actuated
$\text{trace}\left(\mathcal{W}^{-1}(n)\right)$	93.64	676.68	18.16
Average Number of Leaders: d	40	160	200

To have a fair comparison, we normalize the resulting schedule of algorithm 6 such that the sum of the scalings satisfies $\sum_{k=0}^{n-1}\sum_{i=1}^{m} s_i^2(k) = dn$, where $d = 40$. The value of the controllability metric for the materialized result of Algorithm 6 is 18.54, which is much closer to the controllability metric of the fully actuated case.

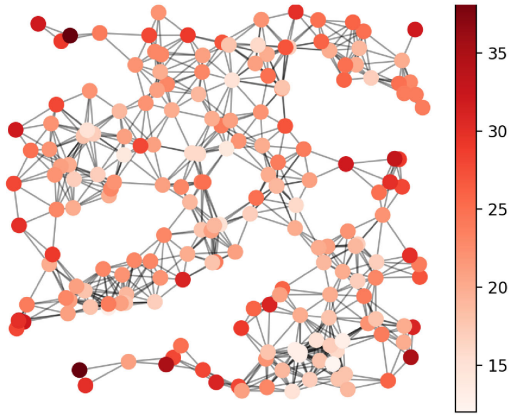


Fig. 3. Dynamical network consists of 200 agents that are randomly distributed in a 1×1 square-shape area in space and are coupled over a proximity graph. Every agent is connected to all of its spatial neighbors within a closed ball of radius $r = 0.125$. Node colors are proportional to the total number of active steps during time steps 0 to 199 from least (white) to greatest (red) based on Algorithm 6 where $d = 40$ (i.e., which means that, on average, only 20% of agents are controlled at each time).

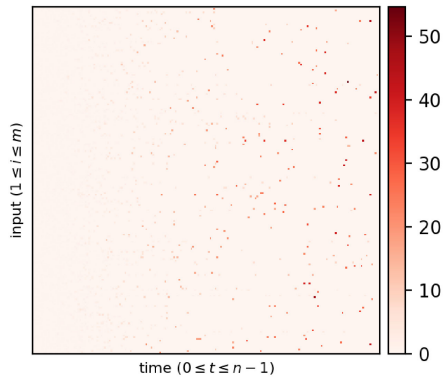


Fig. 4. Sparse schedule based on Algorithm 6 for a given network in Fig. 3 where $d = 40$. This dynamical network has $m = 200$ inputs; however, on average, only 20% are active at each time between 0 to $n-1$. The color of element (i, k) is proportional to the scaling factor $s_i^2(k)$, where $i \in [200]$ and $k+1 \in [200]$.

and for the fully actuated case (i.e., $B = I_8$), we have

$$\text{trace}\left(\sum_{i=0}^{n-1} A^i BB^\top (A^i)^\top\right)^{-1} = 0.132.$$

We compare our method with simple-random and periodical switching methods, which are depicted in Fig. 2, and obtain systemic controllability performances, which are presented in Table II.

Example 2: Let us consider a dynamic network consisting of $n = 200$ agents/nodes, which are randomly distributed in a 1×1 square-shape area in space and are coupled over a proximity graph. Every agent is connected to all of its spatial neighbors within a closed ball of radius $r = 0.125$. Assume that the state-space matrices of this network are given by

$$A = I_n - \frac{1}{n}L, \text{and} B = I_n \tag{49}$$

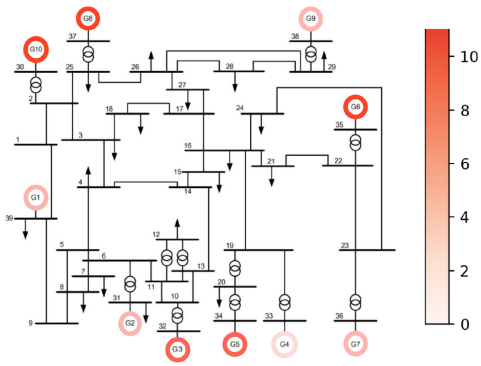


Fig. 5. IEEE 10-generator 39-bus power system network (figure is adapted from [32]). Generator colors are proportional to the total number of active steps during time steps 0 to 19 from least (white) to greatest (red) based on Algorithm 7 where $d = 4$ (i.e., which means that, on average, only four generators are controlled at each time).

where L is the Laplacian matrix of the underlying graph given by Fig. 3 . Now, we consider the actuator scheduling problem discussed in Section V. For undirected consensus networks, a similar problem arises in assignment of a prespecified number of

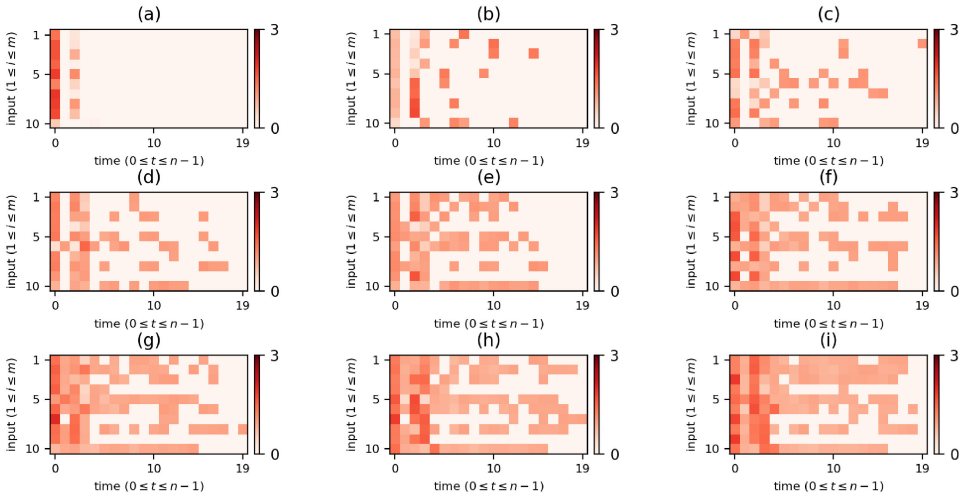


Fig. 6. Subplots (a)–(i) presents nine weighted sparse schedules for Example 3 based on the proposed deterministic method (Algorithm 3), where $d \in \{1.05, 1.75, 2.30, 3.10, 3.95, 4.60, 5.25, 5.75, 6.35\}$ is the average number of active actuators at each time, respectively. The color of element (i, k) is proportional to the scaling factor $s_i^2(k)$ where $i \in [10]$ and $k+1 \in [20]$.

active agents, as leaders, in order to minimize the controllability metric, e.g., the average controllability energy (cf. [48], [49]). In our setup, each leader i in addition to relative information exchange with its neighbors (based on Laplacian matrix L), it also has access to a control input $u_i(.)$. This system is controllable with only a few inputs/leaders;[9] however, the amount of the average control energy with a static actuator/leader schedule is too large even for a large number of leaders (see Table III). On the other hand, with a time-varying strategy, the resulting performance is close to the fully actuated case even with a small number of leaders. Therefore, instead of choosing the same leaders at every time step, we choose/switch leaders over a given time horizon to further decrease the controllability metric.

Fig. 3 shows the underlying graph, and node colors are proportional to the total number of active steps during time steps 0 to 199 from least (white) to greatest (red). Fig. 4 depicts a sparse schedule based on Algorithm 6.

Example 3 (Power Network): The problem is to select a set of generators to be involved in the widearea damping control of power systems. We apply our sparse scheduling approach on the IEEE 39-bus test system (a.k.a. the 10-machine New England Power System; see Fig. 5) [32], [33]. The single line diagram presented in this figure comprises generators (G_i where $i \in [10]$), loads (arrows), transformers (double circles), buses (bold line segments with number $i \in [39]$), and lines between buses (see [32], [33]).

The goal of the widearea damping control is to damp the fluctuations between generators and synchronize all generators. The voltage at each generator is adjusted by the control inputs (e.g., HVdc lines and storages) to regulate the power output.

We start with a model representing the interconnection between subsystems. Consider the swing dynamics

$$m_i\ddot{\theta}_i + d_i\dot{\theta}_i = -\sum_{j\sim i} k_{ij}(\theta_i - \theta_j) + u_i$$

[9]The system is not controllable with only one input, because A does not have distinct eigenvalues [49].

where θ_i is the rotor angle state and $w_i := \dot{\theta}_i$ is the frequency state of generator i. We assume this power grid model consists of $n = 10$ generators [32], [33]. The state-space model of the swing equation used for frequency control in power networks can be written as follows:

$$\begin{bmatrix} \dot{\theta}(t) \\ \dot{w}(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}L & -M^{-1}D \end{bmatrix} \begin{bmatrix} \theta(t) \\ w(t) \end{bmatrix} + \begin{bmatrix} 0 \\ M^{-1} \end{bmatrix} u(t)$$

$$y(t) = \begin{bmatrix} \theta(t) \\ w(t) \end{bmatrix}$$

where M and D are diagonal matrices with inertia coefficients and damping coefficients of generators and their diagonals, respectively.

We assume that both rotor angle and frequency are available for measurement at each generator. This means each subsystem in the power network has a phase measurement unit (PMU). The PMU is a device that measures the electrical waves on an electricity grid using a common time source for synchronization. The system is discretized to the discrete-time LTI system with state matrices A, B, and C and the sampling time of 0.2 second (the matrices are borrowed from [50]).

Fig. 6 depicts nine sparse schedules based on the proposed deterministic method (see Algorithms 3) for different values of d. The sparsity degree of each schedule is captured by d. As d increases the number of nonzero scalings (i.e., activations) increases while the controllability metric decreases (improves). Fig. 7 compares the results of Algorithms 3, 7, 8, and 9. The plot presents the values of the average controllability energy (A-optimality) versus the average number of active actuators. To have a fair comparison, we normalize the resulting schedules of all the methods such that the sum of all the scalings satisfies $\sum_{k=0}^{n-1}\sum_{i=1}^{m} s_i^2(k) = nd$.

As one expects, Algorithms 3, 7, and 9 outperform Algorithm 8. One observes that Algorithms 3 and 7 perform nearly as optimal as the time-varying greedy method Algorithm 9; however, based on our results, we have theoretical guaranteed

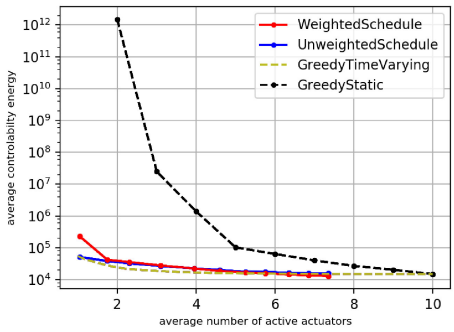


Fig. 7. This plot compares four different methods (Algorithms 3, 7, 8, and 9) for obtaining sparse actuator schedules of the 10-machine New England Power System in Example 3. The plot presents the values of average controllability energy (A-optimality) versus the average number of active actuators at each time (d).

performance bounds for Algorithms 3 and 7. Furthermore, the usefulness of Algorithms 3 and 7 accentuates itself when the number of active actuators on average is not too small; and potentially can result in a better solution compare to Algorithm 9 (see Fig. 7).

VIII. Concluding Remarks

In this article, we have shown how recent advances in matrix reconstruction and graph sparsification literature can be utilized to develop subset selection tools for choosing a relatively small subset of actuators to approximate certain controllability measures. Current approaches based on polynomial time relaxations of the subset selection problem require an extra multiplicative factor of $\log n$ sensors/actuators times the minimal number in order to just maintain controllability/observability. Furthermore, when the control energy is chosen as the cost, submodularity-based approaches fail to guarantee the performance using greedy methods. In contrast, we show that there exists a polynomial-time actuator schedule that on average selects only a constant number of actuators at each time, to approximate controllability measures. Similar results can be developed for the sensor selection problem. A potential future direction is to see whether this approach can be used to develop an efficient scheme for minimal reachability problems.

References

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

[2] 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.

[3] Y.-Y. Liu and A.-L. Barabási, "Control principles of complex systems," *Rev. Mod. Phys.*, vol. 88, no. 3, 2016, Art. no. 035006.

[4] P. V. Chanekar, N. Chopra, and S. Azarm, "Optimal actuator placement for linear systems with limited number of actuators," in *Proc. Amer. Control Conf.*, May 2017, pp. 334–339.

[5] P. Müller and H. Weber, "Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems," *Automatica*, vol. 8, no. 3, pp. 237–246, 1972.

[6] 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.

[7] A. Olshevsky, "Minimum input selection for structural controllability," in *Proc. Amer. Control Conf.*, Jul. 2015, pp. 2218–2223.

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

[9] E. Nozari, F. Pasqualetti, and J. Cortés, "Time-invariant versus time-varying actuator scheduling in complex networks," in *Proc. Amer. Control Conf.*, May 2017, pp. 4995–5000.

[10] A. Yazıcıoğlu, W. Abbas, and M. Egerstedt, "Graph distances and controllability of networks," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 4125–4130, Dec. 2016.

[11] 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.

[12] S. Pequito, S. Kar, and A. P. Aguiar, "On the complexity of the constrained input selection problem for structural linear systems," *Automatica*, vol. 62, pp. 193–199, 2015.

[13] A. Chakrabortty and M. D. Ilić, *Control and Optimization Methods for Electric Smart Grids*, vol. 3. Berlin, Germany: Springer, 2011.

[14] F. A. Chandra, G. Buzi, and J. Doyle, "Glycolytic oscillations and limits on robust efficiency," *Science*, vol. 333, no. 6039, pp. 187–192, 2011.

[15] L. Marucci *et al.*, "How to turn a genetic circuit into a synthetic tunable oscillator, or a bistable switch," *PLoS One*, vol. 4, no. 12, 2009, Art. no. e8083.

[16] I. Rajapakse, M. Groudine, and M. Mesbahi, "What can systems theory of networks offer to biology?" *PLoS Comput. Biol.*, vol. 8, no. 6, 2012, Art. no. e1002543.

[17] M. Siami and J. Skaf, "Structural analysis and optimal design of distributed system throttlers," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 540–547, Feb. 2018.

[18] R. E. Kalman, "Mathematical description of linear dynamical systems," *J. Soc. Ind. Appl. Math., Ser. A, Control*, vol. 1, no. 2, pp. 152–192, 1963.

[19] M. Athans, "On the determination of optimal costly measurement strategies for linear stochastic systems," *Automatica*, vol. 8, no. 4, pp. 397–412, 1972.

[20] A. Jadbabaie, A. Olshevsky, G. J. Pappas, and V. Tzoumas, "Minimal reachability is hard to approximate," *IEEE Trans. Autom. Control*, vol. 64, no. 2, pp. 783–789, Feb. 2019.

[21] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Minimal reachability problems," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 4220–4225.

[22] V. Tzoumas, K. Gatsis, A. Jadbabaie, and G. J. Pappas, "Resilient monotone submodular function maximization," in *Proc. IEEE 56th Annu. Conf. Decision Control*, Melbourne, VIC, Australia, 2017, pp. 1362–1367, doi: 10.1109/CDC.2017.8263844.

[23] O. Kempthorne, *The Design and Analysis of Experiments*. Hoboken, NJ, USA: Wiley, 1952.

[24] Z. Allen-Zhu, Y. Li, A. Singh, and Y. Wang, "Near-optimal design of experiments via regret minimization," in *Proc. 34th Int. Conf. Mach. Learn.*, Aug. 2017, vol. 70, pp. 126–135.

[25] Y. T. Lee and H. Sun, "An SDP-based algorithm for linear-sized spectral sparsification," in *Proc. 49th Annu. ACM Sigact Symp. Theory Comput.*, 2017, pp. 678–687.

[26] D. A. Spielman and N. Srivastava, "Graph sparsification by effective resistances," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, New York, NY, USA, 2008, pp. 563–568.

[27] J. Batson, D. A. Spielman, and N. Srivastava, "Twice-ramanujan sparsifiers," *SIAM J. Comput.*, vol. 41, no. 6, pp. 1704–1721, 2012.

[28] M. Rudelson and R. Vershynin, "Sampling from large matrices: An approach through geometric functional analysis," *J. ACM*, vol. 54, no. 4, Jul. 2007, Art. no 21.

[29] A. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing families II: Mixed characteristic polynomials and the Kadison—Singer problem," *Annu. Math.*, pp. 327–350, 2015.

[30] N. Srivastava and L. Trevisan, "An Alon–Boppana type bound for weighted graphs and lowerbounds for spectral sparsification," in *Proc. 29th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2018, pp. 1306–1315.

[31] M. A. Elizondo *et al.*, "Interarea oscillation damping control using high-voltage dc transmission: A survey," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6915–6923, Nov. 2018.

[32] I. E. Atawi, "An advance distributed control design for wide-area power system stability," Ph.D. dissertation, Dept. Elect. Eng., Univ. Pittsburgh, Pittsburgh, PA, USA, 2013.

[33] Z. Liu *et al.*, "Minimal input selection for robust control," in *Proc. IEEE 56th Annu. Conf. Decision Control*, 2017, pp. 2659–2966.

[34] S. Boyd, "Lecture notes for EE263," *Introduction to Linear Dynamical Systems*, 2008. [Online]. Available: http://ee263.stanford.edu/archive/ee263_course_reader.pdf

[35] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Syst. Mag.*, vol. 36, no. 1, pp. 26–46, Feb. 2016.
[36] M. Siami and A. Jadbabaie, "Deterministic polynomial-time actuator scheduling with guaranteed performance," in *Proc. Eur. Control Conf.*, 2018, pp. 113–118.
[37] A. Jadbabaie, A. Olshevsky, and M. Siami, "Limitations and tradeoffs in minimum input selection problems," in *Proc. Annu. Amer. Control Conf.*, 2018, pp. 185–190.
[38] C. Boutsidis, P. Drineas, and M. Magdon-Ismail, "Near-optimal column-based matrix reconstruction," *SIAM J. Comput.*, vol. 43, no. 2, pp. 687–717, 2014.
[39] J. W. Silverstein, "The stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices," *Random Matrix Theory and Its Applications. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap*, vol. 18, pp. 1–25, 2009.
[40] M. A. Camacho, "Spectral sparsification: The barrier method and its applications," Ph.D. dissertation, Harvard College, 2014. [Online]. Available: http://nrs.harvard.edu/urn-3:HUL.InstRepos:12553868
[41] M. Siami and N. Motee, "Network abstraction with guaranteed performance bounds," *IEEE Trans. Autom. Control*, vol. 63, no. 10, pp. 3301–3316, Oct. 2018.
[42] M. Siami and N. Motee, "Systemic measures for performance and robustness of large–scale interconnected dynamical networks," in *Proc. 53rd IEEE Conf. Decis. Control*, Dec. 2014, pp. 5119–5124.
[43] M. Siami and N. Motee, "Growing linear dynamical networks endowed by spectral systemic performance measures," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 2091–2106, Jul. 2018.
[44] S. N. Ravi, V. Ithapu, S. Johnson, and V. Singh, "Experimental design on a budget for sparse linear models and applications," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 583–592.
[45] F. Pukelsheim, *Optimal Design of Experiments*, vol. 50. Philadelphia, PA, USA: SIAM, 1993.
[46] D. A. Spielman, "Lecture notes in spectral and algebraic graph theory," Yale University, New Haven, CT, USA, Nov. 14, 2018. [Online]. Available: http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf
[47] A. Olshevsky, "On (non)supermodularity of average control energy," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1177–1181, Sep. 2018.
[48] F. Lin, M. Fardad, and M. R. Jovanović, "Algorithms for leader selection in stochastically forced consensus networks," *IEEE Trans. Autom. Control*, vol. 59, no. 7, pp. 1789–1802, Jul. 2014.
[49] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt, "Controllability of multi-agent systems from a graph-theoretic perspective," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 162–186, 2009.
[50] G. Fazelnia, R. Madani, A. Kalbat, and J. Lavaei, "Convex relaxation for optimal distributed control problems," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 206–221, Jan. 2017.

Milad Siami (Member, IEEE) received the dual B.Sc. degrees in electrical engineering and pure mathematics and the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2009 and 2011, respectively, and the M.Sc. and Ph.D. degrees in mechanical engineering from Lehigh University, Bethlehem, PA, USA, in 2014 and 2017, respectively.

From 2009 to 2010, he was a Research Student with the Department of Mechanical and Environmental Informatics, the Tokyo Institute of Technology, Tokyo, Japan. He was a Postdoctoral Associate with the Institute for Data, Systems, and Society, MIT, from 2017 to 2019. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. His research interests include distributed control systems, distributed optimization, and applications of fractional calculus in engineering.

Dr. Siami received a Gold Medal at National Mathematics Olympiad, Iran in 2003 and the Best Student Paper Award at the 5th IFAC Workshop on Distributed Estimation and Control in Networked Systems in 2015. He was awarded RCEAS Fellowship in 2012, Byllesby Fellowship in 2013, Rossin College Doctoral Fellowship in 2015, and Graduate Student Merit Award in 2016 at Lehigh University.

Alexander Olshevsky (Member, IEEE) received the B.S. degree in applied mathematics and in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, both in 2004, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2006 and 2010, respectively.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA. Prior to this position, he was a Faculty Member with the University of Illinois at Urbana-Champaign, IL, USA. He was a Postdoctoral Scholar with the Department of Mechanical and Aerospace Engineering, Princeton University, from 2010 to 2012 before joining the University of Illinois at Urbana-Champaign in 2012. His research interests include control theory, optimization, and machine learning, especially in distributed, networked, and multiagent settings.

Dr. Olshevsky received the NSF CAREER Award, the Air Force Young Investigator Award, the ICS Prize from INFORMS for Best Paper on the interface of operations research and computer science, and a SIAM Paper Prize for annual paper from the SIAM Journal on Control and Optimization chosen to be reprinted in SIAM Review, a Best Paper Award in 2019 from the International Medical Informatics Association on Clinical Research Informatics.

Ali Jadbabaie (Fellow, IEEE) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1995 the M.S. degree in electrical and computer engineering from the University of New Mexico, Albuquerque, NM, USA, in 1997 and the Ph.D. degree in control and dynamical systems from California Institute of Technology, Pasadena, CA, USA, in 2000.

He is the JR East Professor of engineering, the Associate Director of the Institute for Data, Systems and Society, and the Director of the Sociotechnical Systems Research Center, MIT, Cambridge, MA, USA. He holds faculty appointments with the Department of Civil and Environmental Engineering and is a Principal Investigator with the Laboratory for Information and Decision Systems. He was a Postdoctoral Scholar with Yale University before joining the faculty at Penn in July 2002. Prior to joining MIT, he was the Alfred Fitler Moore Professor of network science and held secondary appointments in computer and information science and operations, information, and decisions in the Wharton School. His current research interests include the interplay of dynamic systems and networks with specific emphasis on multiagent coordination and control, distributed optimization, network science, and network economics.

Prof. Jadbabaie is the Inaugural Editor-in-Chief of the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING and an Associate Editor of the Informs Journal Operations Research. He received the National Science Foundation Career Award, an Office of Naval Research Young Investigator Award, the O. Hugo Schuck Best Paper Award from the American Automatic Control Council, the George S. Axelby Best Paper Award from the IEEE Control Systems Society, and the 2016 Vannevar Bush Fellowship from the office of Secretary of Defense.

Non-asymptotic Concentration Rates in Cooperative Learning Part II: Inference on Compact Hypothesis Sets

César A. Uribe, Alex Olshevsky, and Angelia Nedić

***Abstract*— We study the problem of cooperative inference where a group of agents interacts over a network and seeks to estimate a joint parameter that best explains a set of network-wide observations using local information only. Agents do not know the network topology or the observations of other agents. We explore a variational interpretation of the Bayesian posterior and its relation to the stochastic mirror descent algorithm to prove that, under appropriate assumptions, the beliefs generated by the proposed algorithm concentrate around the true parameter exponentially fast. In Part I of this two-part paper series, we focus on providing a variation approach to distributed Bayesian filtering. Moreover, we develop explicit and computationally efficient algorithms for observation models in the exponential families. Additionally, we provide a novel non-asymptotic belief concentration analysis for distributed non-Bayesian learning on finite hypothesis sets. This new analysis method is the basis for the results presented in Part II. Part II provides the first non-asymptotic belief concentration rate analysis for distributed non-Bayesian learning over networks on compact hypothesis sets. Additionally, we provide extensive numerical analysis for various distributed inference tasks on networks for observational models in the exponential family of distributions.**

***Index Terms*— Distributed Inference, non-Bayesian social learning, estimation over networks, non-asymptotic rates.**

I. Introduction

The increasing amount of data generated by recent applications of distributed systems such as social media, sensor networks, and cloud-based databases has brought considerable attention to distributed data processing, in particular the design of distributed algorithms that take into account the communication constraints and make collective decisions in a distributed manner [1]–[11]. In a distributed system, interactions between agents are usually constrained by the network structure, and agents can only use locally available information. This contrasts with centralized approaches where all information and computation resources are available at a single location [12]–[15].

One traditional problem in decision-making is that of parameter estimation. Given a set of noisy observations coming from a joint distribution, one would like to estimate a parameter or distribution that minimizes a specific loss function. For example, Maximum a Posteriori (MAP) or Minimum Least Squared Error (MLSE) estimators fit a parameter to some model of the observations. Both MAP and MLSE estimators require some form of Bayesian posterior computation based on models that explain the observations for a given parameter. Computation of such a posteriori distributions depends on having exact models about the likelihood of the corresponding observations. This is one of the main difficulties of using Bayesian approaches in a distributed setting. A fully Bayesian approach is not possible because full knowledge of the network structure or other agents' likelihood models may not be available [16]–[18].

Following the seminal work of Jadbabaie et al. in [1], [19], [20], there have been many studies of distributed non-Bayesian update rules over networks. In this case, agents are assumed to be boundedly rational (i.e., they fail to aggregate information in a fully Bayesian way [21]). Proposed non-Bayesian algorithms involve an aggregation step, typically consisting of weighted geometric or arithmetic average of the received beliefs [7], [22]–[25], and a Bayesian update with the locally available data [18], [26]. Lalitha et al. [27], Qipeng et al. [28], [29], Shahrampour et al. [20], [30], [31], and Rahimian et al. [32] have proposed variations of the non-Bayesian approach and proved consistent, geometric and non-asymptotic convergence rates for a general class of distributed algorithms; from asymptotic analysis to non-asymptotic bounds [33], [34], to time-varying directed graphs [35]. Su et al. [36] have also considered adversarial agents and transmission and node failures. Constant elasticity of substitution models [37], minimum operators [38], [39], and uncertain models [40] have been also studied. See [41] and [42] for an extended literature review.

We build upon the work in [43] on non-asymptotic behaviors of Bayesian estimators to derive new non-asymptotic concentration results for distributed learning algorithms. In contrast to the current results, which assume a finite hypothesis set, we extend the framework to compact sets of hypotheses in this paper. Our results show that, in general, the network

C.A. Uribe is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77006 USA e-mail: cauribe@rice.edu.

A. Olshevsky is with the Department of ECE and Division of Systems Engineering, Boston University, Boston, MA, 02215 USA e-mail: alexols@bu.edu.

A. Nedić is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, 85287 USA e-mail: angelia.nedich@asu.edu.

structure will induce a transient time after which all agents learn at an independent network rate, and this rate is geometric.

The main contribution of this paper (Part II) are as follows:

- We provide the first non-asymptotic belief concentration analysis for non-Bayesian distributed learning over ***compact hypothesis sets***.
- We show the proposed update rule concentrates its beliefs on compact balls around the optimal set at a geometric rate.
- We provide extensive numerical results for various distributed inference tasks with observational models in the exponential family of distributions.

The rest of this paper is organized as follows. Section II introduces the problem setup, and it describes the networked observation model and the inference task. Section III shows our main results about the exponential concentration of beliefs around the true parameter. Section III presents the non-asymptotic concentration analysis for the case when the hypothesis set is a compact subset of $\mathbb{R}^d$. Section IV presents a set of numerical analysis and simulation results for the proposed algorithms for the distributed estimation of parameters of distributions from the exponential family for various networks topologies and number of agents. Finally, conclusions, open problems, and potential future work are discussed.

Notation: Random variables are denoted with upper-case letters, e.g., X, while the corresponding lower-case are used for their realizations, e.g., x. Subscripts denote time indices, and the letter k or t is generally used. Superscripts denote agent indices, and the letters i or j are used. We write $[A]_{ij}$ or a_{ij} to denote the entry of a matrix A in its i-th row and j-th column. We use A' for the transpose of a matrix A and x' for the transpose of a vector x. The complement of a set B is denoted as B^c.

II. Problem Setup

We introduce the learning problem from a centralized perspective, where all information is available at a single location. Later, we will generalize the setup to the distributed setting where only partial and distributed information is available.

Assume that we observe a sequence of independent random variables $X_1, X_2, \ldots$, all taking values in some measurable space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{X}$ is the realization space and $\mathcal{A}$ is the corresponding σ-algebra. The random variables $\{X_i\}$ are assumed identically distributed with a common *unknown* distribution P on $\mathcal{X}$, i.e., $X_k \sim P$ for all k. In addition, we have a statistical model $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ composed by a parametrized family of probability measures on the sample space $(\mathcal{X}, \mathcal{A})$, where the map $\Theta \to \mathscr{P}$ from parameter to distribution is injective. Moreover, all distributions in the model are dominated[1] by a σ-finite measure λ[2], with corresponding densities $p_\theta = dP_\theta/d\lambda$. Assume also that the model $\mathscr{P}$ is well-specified, thus *there exists a* θ^* *such that* $P_{\theta^*} = P$. The objective is to estimate θ^* based on the sequence of received observations $x_1, x_2, \ldots$. For example, given a random variable X, the maximum likelihood estimator (MLE) is

$$\hat{\theta}(X) = \arg\max_{\theta\in\Theta} p_\theta(X) = \arg\max_{P\in\mathscr{P}} p(X).$$

Following a Bayesian approach, the parameter is represented as a random variable ϑ on the set Θ equipped with a σ-algebra $\mathcal{T}$ and a prior probability measure μ_0 on the measurable space $(\Theta, \mathcal{T})$, where $\mathcal{T}$ is countably generated. Moreover, we assume the existence of a probability measure Π on the product space $(\mathcal{X} \times \Theta)$ with σ-algebra $(\mathcal{A} \times \mathcal{T})$. Furthermore, the densities $p_\theta(x)$ are measurable functions of θ for any $x \in \mathcal{X}$. We define μ_k as the posterior distribution given the sequence of observations up to time k, i.e.,

$$\mu_k(B) = \Pi(\vartheta \in B \mid X_1, \ldots, X_k) = \frac{\int_B \prod_{t=1}^k p_\theta(X_t) d\mu_0(\theta)}{\int_\Theta \prod_{t=1}^k p_\theta(X_t) d\mu_0(\theta)}, \quad (1)$$

for all $B \in \mathcal{T}$ (note that we used the independence of the observations at each time step).

Assuming that all observations, up to time k, are readily available at a centralized location, under appropriate conditions, the recursive Bayesian posterior in (1) will be consistent in the sense that the beliefs μ_k will concentrate around θ^*; see [44], [45], and [46] for a formal statement. Furthermore, several authors have studied the rate at which this concentration occurs, in both asymptotic and non-asymptotic regimes [43], [47], [48].

Now consider the case where there is a network of n agents observing the process $X_1, X_2, \ldots$, where X_k is now a random vector belonging to the product space $\prod_{i=1}^n \mathcal{X}^i$ and $X_k = [X_k^1, X_k^2, \ldots, X_k^n]'$. Specifically, agent i observes the sequence $X_1^i, X_2^i, \ldots$, where X_k^i is now distributed according to an unknown distribution P^i, effectively making $X_k \sim \boldsymbol{P} = \prod_{i=1}^n P^i$. The statistical model is now distributed, where each agent agent i has a private family of distributions $\mathscr{P}^i = \{P_\theta^i : \theta \in \Theta\}$ it would like to fit to the observations. However, the goal is for *all* agents to agree on a *single* θ that best explains the complete set of observations instead of their local observations only. In other words, the agents collaboratively seek to find θ^* such that $\boldsymbol{P}_{\theta^*} = \prod_{i=1}^n P_{\theta^*}^i = \prod_{i=1}^n P^i = \boldsymbol{P}$.

Agents interact over a network defined by an undirected graph $\mathcal{G} = (V, E)$, where $V = \{1, 2, \ldots, n\}$ is the set of agents and E is a set of undirected edges, i.e., $(i, j) \in E$ if and only if agents i and j can communicate with each other. We study a simple interaction model where, at each step, agents exchange their beliefs with their neighbors in the graph. Thus at every time step k, agent i will receive the sample x_k^i from X_k^i as well as the beliefs of its neighboring agents, i.e., it will receive μ_{k-1}^j for all j such that $(i, j) \in E$. We assume agents are oblivious to the network topology and the private family of distributions of other agents. Thus, fully Bayesian approaches cannot be used. *Our goal is to design a learning procedure that is both distributed and consistent. That is, we are interested in a belief update algorithm that aggregates information in a non-Bayesian manner and guarantees that the beliefs of all agents will concentrate around* θ^*.

As shown in Part I of this paper series [49], the above

[1] A measure μ is dominated by (or absolutely continuous with respect to) a measure λ if $\lambda(B) = 0$ implies $\mu(B) = 0$ for every measurable set B.

[2] A positive measure defined on a σ-algebra of subsets of a set X is called σ-finite if X is the countable union of measurable sets with finite measure.

problem can be written as the optimization problem

$$\min_{\theta\in\Theta} F(\theta) \triangleq D_{KL}\left(\boldsymbol{P}\|\boldsymbol{P}_\theta\right) = \sum_{i=1}^{n} D_{KL}(P^i\|P^i_\theta). \tag{2}$$

We propose the following algorithm as a distributed version of the stochastic mirror descent for the solution of (2):

$$d\mu^i_{k+1} = \arg\min_{\pi\in\Delta_\Theta}\left\{\langle -\log p^i_\theta(x^i_{k+1}),\pi\rangle + \sum_{j=1}^{n} a_{ij} D_{KL}(\pi\|d\mu^j_k)\right\}$$
$$\text{where } \theta\sim\pi, \tag{3}$$

with $a_{ij}>0$ denoting the weight that agent i assigns to beliefs coming from its neighbor j. Specifically, $a_{ij}>0$ if $(i,j)\in E$ or $j=i$, and $a_{ij}=0$ if $(i,j)\notin E$. Problem (3) has a closed form solution. In particular, the posterior density at each $\theta\in\Theta$ is given by

$$d\mu^i_{k+1}(\theta) \propto p^i_\theta(x^i_{k+1}) \prod_{j=1}^{n} (d\mu^j_k(\theta))^{a_{ij}}, \tag{4}$$

or equivalently, the belief on a measurable set B of an agent i at time $k+1$ is

$$\mu^i_{k+1}(B) \propto \int_B p^i_\theta(x^i_{k+1}) \prod_{j=1}^{n} (d\mu^j_k(\theta))^{a_{ij}}. \tag{5}$$

The update (5) can be viewed as a two-step process: first, every agent constructs an aggregate belief using a weighted geometric average of its own belief and the beliefs of its neighbors, and then each agent performs a Bayes' update using the aggregated belief as a prior. We note that similar arguments in the context of distributed optimization have been proposed in [50], [51] for general Bregman distances. In the case when the number of hypotheses is finite, variations on this update rule were previously analyzed in [27], [30], [33].

III. Belief Concentration Rates

We now turn to the presentation of our main results about the rate at which beliefs generated by (5) concentrate around the true parameter θ^*. In Part I of this paper series [52], we will focus on the case when Θ is a finite set and prove a concentration rate on the beliefs on a Hellinger ball around the optimal hypothesis. Contrary to Part I [52], in this section, we focus on the case when Θ is a compact subset of $\mathbb{R}^d$. Our proof techniques use concentration arguments for beliefs on Hellinger balls from the recent work in [43] which, in turn, builds on the classic paper of [53].

We begin with two subsections focusing on background information, definitions, and assumptions.

A. Background: Hellinger Distance and Coverings

The *squared* Hellinger distance between two probability distributions P and Q is given by,

$$h^2(P,Q) = \frac{1}{2}\int\left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda, \tag{6}$$

where P and Q are dominated by λ. Moreover, the Hellinger distance satisfies the property that $0\le h(P,Q)\le 1$.

We equip the set of all probability distributions $\mathscr{P}$ over the parameter set with the Hellinger distance to obtain the *metric* space $(\mathscr{P},h)$. The metric space induces a topology, where we can define an open ball $\mathcal{B}_r(\theta)$ with a radius $r\in(0,1)$ centered at a point $\theta\in\Theta$, which we use to construct a special covering of subsets $B\subset\mathscr{P}$. Recall that (6) defines the squared Hellinger distance h^2, rather than h.

Definition 1: Define an n-Hellinger ball of radius r centered at θ as

$$\mathcal{B}_r(\theta) = \left\{\hat\theta\in\Theta \,\middle|\, \sqrt{\sum_{i=1}^{n} h^2\left(P^i_\theta, P^i_{\hat\theta}\right)} \le r\right\}.$$

Additionally, when no center is specified, it should be assumed that it refers to θ^*, i.e. $\mathcal{B}_r = \mathcal{B}_r(\theta^*)$.

Given an n-Hellinger ball of radius r, we will use the following notation for a covering of its complement $\mathcal{B}^c_r$. Specifically, we are going to express $\mathcal{B}^c_r$ as the union of finite disjoint and concentric annuli. Let $r\in(0,\sqrt{n})$ and $\{r_l, l=1,\dots,L\}$ be a finite non-increasing sequence such that $r_1=\sqrt{n}$ and $r_L=r$ and express the set $\mathcal{B}^c_r$ as the union of annuli generated by the sequence $\{r_l\}$ as $\mathcal{B}^c_r = \bigcup_{l=1}^{L-1}\mathcal{F}_l$, where $\mathcal{F}_l = \mathcal{B}_{r_l}\setminus\mathcal{B}_{r_{l+1}}$.

B. Background: Assumptions on the Network and Mixing Weights

Naturally, we need some assumptions on the matrix A. For one thing, the matrix A has to be "compatible" with the underlying graph, in that information from node i should not affect node j if there is no edge from i to j in $\mathcal{G}$. At the other extreme, we want to rule out the possibility that A is the identity matrix, which in terms of (5) means nodes do not talk to their neighbors. Formally, we make the following assumption.

Assumption 1: The graph $\mathcal{G}$ and matrix A are such that:

(a) A is symmetric and row stochastic with $[A]_{ij}=a_{ij}>0$ for $i\neq j$ if and only if $(i,j)\in E$.
(b) A has positive diagonal entries, $a_{ii}>0$ for all $i\in V$.
(c) The graph $\mathcal{G}$ is connected.

Assumption 1 is common in the distributed optimization literature. The construction of a set of weights satisfying Assumption 1 can be done in a distributed way, for example, by choosing the so-called "lazy Metropolis" matrix, which is a stochastic matrix given by

$$a_{ij} = \begin{cases} \frac{1}{2\max\{d^i+1,d^j+1\}} & \text{if } (i,j)\in E,\\ 0 & \text{if } (i,j)\notin E,\end{cases}$$

where d^i is the degree (the number of neighbors) of node i. Note that although the above formula only gives the off-diagonal entries of A, it uniquely defines the entire matrix (the diagonal elements are uniquely defined via the stochasticity of A). To choose the weights corresponding to a lazy Metropolis matrix, agents will need to spend an additional round at the beginning of the algorithm broadcasting their degrees to their neighbors.

Assumption 1 can be seen to guarantee that $A^k\to(1/n)\mathbf{1}\mathbf{1}^T$ where $\mathbf{1}$ is the vector of all ones. We will use

the following result based on [30] and [33], that provides convergence rate for the difference $|A^k - (1/n)\mathbf{1}\mathbf{1}^T|$,:

Lemma 1: Let Assumption 1 hold, then the matrix A satisfies the following relation:

$$\sum_{t=1}^{k}\sum_{j=1}^{n}\left|\left[A^{k-t}\right]_{ij} - \frac{1}{n}\right| \leq \frac{4\log n}{1-\delta} \qquad \text{for } i = 1, \ldots, n,$$

where $\delta = 1 - \eta/4n^2$ with η being the smallest positive entry of the matrix A. Furthermore, if A is a lazy Metropolis matrix associated with the graph $\mathcal{G}$, then $\delta = 1 - 1/\mathcal{O}(n^2)$.

See Table II in the appendix for a detailed account of network dependencies with respect to the number of nodes for large classes of networks. The mixing times, in turn, serve as bounds for the spectral gap of the mixing matrices.

C. A Concentration Result for a Compact Set of Hypotheses

Next, we will study the non-asymptotic belief concentration process when the hypothesis set Θ is a compact subset of $\mathbb{R}^d$. We additionally require the map from Θ to $\prod_{i=1}^n P_\theta^i$ to be continuous (where the topology on the space of distributions comes from the Hellinger metric). This will be useful in defining coverings, which will be made clear shortly.

Definition 2: Let (M, d) be a metric space. A subset $S \subseteq M$ is called ε-separated with $\varepsilon > 0$ if $d(x, y) \geq \varepsilon$ for any $x, y \in S$. Moreover, for a set $B \subseteq M$, let $N_B(\varepsilon)$ be the smallest number of d-balls with centers in S of radius ε needed to cover the set B, i.e., such that $B \subseteq \bigcup_{m\in S} \mathcal{B}_\varepsilon(m)$.

Given a decreasing sequence $\sqrt{n} = r_1 \geq r_2 \geq \cdots \geq r_L = r$, we will define the annulus $\mathcal{F}_l$ to be $\mathcal{F}_l = \mathcal{B}_{r_l} \setminus \mathcal{B}_{r_{l+1}}$. Furthermore, S_{ε_l} will denote a maximal ε_l-separated subset of $\mathcal{F}_l$. Finally, $K_l = |S_{\varepsilon_l}|$. As a consequence of our assumption that the map from Θ to $\prod_{i=1}^n P_\theta^i$ is continuous, we have that each K_l is finite (since the image of a compact set under a continuous map is compact).

Remark 1: Note that Definition 2 induces a covering of the sets $\mathcal{F}_l$ by K_l balls of radius ε_l, centered at points $m \in S_{\varepsilon_l}$, i.e., $\mathcal{B}_{\varepsilon_l}(m)$. From this covering we can deduce a partition $\mathcal{F}_l = \bigcup_{m\in S} \mathcal{F}_{m,l}$, where each $\mathcal{F}_{m,l} \subseteq \mathcal{B}_{\varepsilon_l}(m)$ is the intersection of a ball centered at an element in S_{ε_l} with $\mathcal{F}_l$. Thus, we have the following finite covering of $\mathcal{B}_r^c$:

$$\mathcal{B}_r^c \subseteq \bigcup_{l=1}^{L-1} \bigcup_{m\in S_{\varepsilon_l}} \mathcal{F}_{l,m}. \tag{7}$$

Figure 1 shows the elements of a covering for a set $\mathcal{B}_r^c$. The cluster of circles at the top right corner represents the balls $\mathcal{B}_{\varepsilon_l}$ and, for a specific case in the left of the image, we illustrate the set $\mathcal{F}_{l,m}$ shaded in gray.

We will make the following technical assumption that will be convenient for the analysis of the concentration of beliefs on compact sets. In particular, such an assumption holds for the exponential family of distributions.

We will require a continuity assumption of the likelihood with respect to the parameter space Θ for our non-asymptotic analysis.

Assumption 2: The likelihood function $p_\theta(x)$ is continuous on Θ with respect to θ for any $x \in \mathcal{X}$.

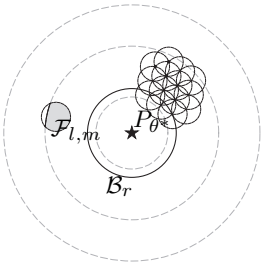


Fig. 1: Creating a covering for a set $\mathcal{B}_r$. ★ represents the correct hypothesis $\boldsymbol{P}_{\theta^*}$. The set $\mathcal{F}_{l,m}$ shaded in gray.

Assumption 2 will hold for large classes of likelihood functions, in particular for the exponential family of distributions. For example, it trivially holds for Gaussian distributions with known variance and known mean. In general, this assumption forbids arbitrarily large changes in the likelihood model for infinitesimal changes in the parameter. We will use this assumption later to guarantee the existence of a parameter inside a closed ball in the parameter space that minimizes the integral likelihood model defined on the ball for any measurable subset of the observation space. Assumption 2 is only a sufficient condition. The interested reader can see [54, Chapter 5] for an extensive account of weaker assumptions to guarantee the congruence of an estimator. Moreover, Assumption 2 will allow us to state the following auxiliary result.

Proposition 2: Let B be a closed n-Hellinger ball (c.f. Definition 1) centered at $\theta_B \in \Theta$, and let Assumption 2 hold for the family of distributions $\mathscr{P}^i = \{P_\theta^i : \theta \in \Theta\}$ for $i \in V$. Then, for every sequence x_t^i for $i \in V$ and $t = 1, \cdots k$ there exists a θ such that:

$$\frac{1}{\mu_0(B)} \int_B \prod_{t=1}^{k}\prod_{j=1}^{n} p_{\hat{\theta}}^j(x_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\hat{\theta}) \geq \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(x_t^j)^{[A^{k-t}]_{ij}}.$$

Proof: The closedness of the n-Hellinger ball B, and the continuity with respect to θ in Assumption 2 are sufficient for extreme values to exist by the Weierstrass extreme value theorem [55, Section 3.4]. ∎

We next provide a concentration result for the logarithmic likelihood of a ratio of densities, which will serve the same technical function as Lemma 4 in Part I of this paper series [52]. We begin by defining two measures. For a hypothesis θ and a measurable set $B \subseteq \Theta$, let $\boldsymbol{P}_B^{\otimes k}$ be the probability distribution with density, (i.e., Radon-Nikodym derivative with respect to $\lambda^{\otimes nk}$),

$$g_B(\boldsymbol{x}^k) = \frac{1}{\mu_0(B)} \int_B \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(x_t^j) d\mu_0(\theta). \tag{8}$$

Similarly, let $\bar{\boldsymbol{P}}_B^{\otimes k}$ be the measure with density

$$\bar{g}_B^i(\boldsymbol{x}^k) = \frac{1}{\mu_0(B)} \int_B \prod_{t=1}^{k}\prod_{j=1}^{n} (p_\theta^j(x_t^j))^{[A^{k-t}]_{ij}} d\mu_0(\theta). \tag{9}$$

Moreover, with some notation abuse define

$$g_\theta(\boldsymbol{x}^k) = \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(x_t^j), \tag{10}$$

$$\bar{g}_\theta^i(\boldsymbol{x}^k) = \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(x_t^j)^{[A^{k-t}]_{ij}}. \tag{11}$$

Note that $\bar{\boldsymbol{P}}_B^{\otimes k}$'s are not probability distributions due to the exponential weights. Nonetheless, they are bounded and non-negative. The next lemma shows the concentration of the logarithmic ratio of a weighted density as defined in (9) for a set B and a density at an arbitrary hypothesis $\hat{\theta} \in \Theta$, in terms of the probability distribution $\boldsymbol{P}_{\hat{\theta}}^{\otimes k}$.

Lemma 3: Let Assumptions 1 and 2 hold. Consider a measurable set $B \subset \Theta$ with a positive measure, and assume that $B \subset \mathcal{B}_r(\theta_B)$ where $\theta_B \in \Theta$. Moreover, let $\theta \in \Theta / \mathcal{B}_r(\theta_B)$ be an element of the parameter space not in B. Then, for all $y \in \mathbb{R}$, and all $i \in V$,

$$\mathbb{P}_\theta\left[\log \frac{\bar{g}_B^i(\boldsymbol{X}^k)}{\bar{g}_\theta^i(\boldsymbol{X}^k)} \geq y\right] \leq \exp\left(-\frac{y}{2} + \frac{4\log n}{1-\delta} - \frac{k}{n}\left(\sqrt{\sum_{j=1}^{n} h^2\big(P_{\theta_B}^j, P_\theta^j\big)} - r\right)^2\right),$$

where $\mathbb{P}_\theta$ is the probability measure that gives $\boldsymbol{X}^k$ a distribution $\boldsymbol{P}_\theta^{\otimes k}$ with density g_θ as defined in (10).

The proof of Lemma 3 can be found in the Appendix. Lemma 3 provides a concentration result for the logarithmic ratio between weighted densities over a subsets B and a density on an arbitrary point θ. The terms involving the auxiliary variable y and the influence of the graph, via δ are the same as in Lemma 4 in Part I of this paper series [52]. Moreover, the rate at which this bound decays exponentially is influenced by the radius of the Hellinger balls B and θ.

Next, we state a technical assumption about the concentration of the initial beliefs around the optimal set of hypotheses

Assumption 3: Let $\mathcal{B}_{r_1}$ and $\mathcal{B}_{r_2}$ two n-Hellinger balls centered at θ^* with radius r_1 and r_2 respectively with $r_1 \leq r_2$. Then, there exists a positive constant $c_{n,d}$ possibly depending on n and d, such that the following property holds for the prior distribution:$\mu_0(\mathcal{B}_{r_2}) \leq \exp\left(c_{n,d} r_2 / r_1\right) \mu_0(\mathcal{B}_{r_1})$.

Assumption 3 states that the initial beliefs should assign sufficient mass on the balls around the optimal hypothesis. In particular, when we select r_2 such that $\mu_0(\mathcal{B}_{r_2}) = 1$, then $\mu_0(\mathcal{B}_{r_1}) \geq \exp\left(-c_0/r_1\right)$, imposing a minimum mass on the n-Hellinger ball of radius r_1 around θ^*. Assumption 3 is analogous to [43, Assumption 2] where a similar condition is required. Moreover, as stated in [43] such a condition is satisfied by many parametric models defined over bounded sets of $\mathbb{R}^d$ and the uniform initial beliefs.

We are ready now to state our main result regarding the concentration of beliefs around θ^* for compact sets of hypotheses.

Theorem 4: Let Assumptions 1, 3, and 2 hold, and let $\sigma \in (0,1)$ be a given probability tolerance level, $r \in (0, \sqrt{n})$, and $\{r_l, l = 1, \cdots, L\}$ be a finite strictly decreasing sequence such that $r_1 = \sqrt{n}$, $r_L = r$, and $r_l \leq 2r_{l+1}$. Moreover, assume initial beliefs of agents are equal to each other almost surely. Then, the beliefs $\{\mu_k^i\}$, $i \in V$, generated by Eq. (5) have the following property: with probability $1 - \sigma$,

$$\mu_k^i(\mathcal{B}_r) \geq 1 - \sum_{l=1}^{L-1} \exp\left(-\frac{k}{16n} r_{l+1}^2\right), \quad \forall i \in V, k \geq N,$$

where

$$N \geq \inf\left\{t \geq 1 \middle| \sum_{l=1}^{L-1} K_l \exp\left(\frac{4\log n}{1-\delta} - \frac{t}{16n} r_{l+1}^2\right) < \frac{\sigma}{2}, \text{ and } t \geq \left(64\sqrt{2} c_{n,d} n/(\sigma r_{l+1})\right)^2\right\}.$$

K_l is the number of balls of radius $r_{l+1}/32$ required to cover the annulus $\mathcal{F}_l = \mathcal{B}_{r_l} \setminus \mathcal{B}_{r_{l+1}}$, and $\delta = 1 - \eta/n^2$, where η is the smallest positive element of the matrix A.

Proof: Following similar arguments as the proof of Theorem 5 in Part I of this paper series, lets start by analyzing the evolution of the beliefs on a measurable set B with $\theta^* \in B$. From (5) we have that

$$\mu_k^i(B) = \frac{\int_B \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} \prod_{j=1}^{n} d\mu_0^j(\theta)^{[A^k]_{ij}}}{\int_\Theta \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} \prod_{j=1}^{n} d\mu_0^j(\theta)^{[A^k]_{ij}}} \geq 1 - \frac{\int_{B^c} \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\theta)}{\int_B \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\theta)}$$

where $\prod_{j=1}^{n} d\mu_0^j(\theta)^{[A^k]_{ij}} = d\mu_0(\theta)$ follows from the assumption of equal initial beliefs almost surely for all agents.

Now lets focus specifically on the case where B is a n-Hellinger ball of radius $r \in (0,1)$ with center at θ^*, i.e., $\mathcal{B}_r$. For analysis purposes, we will let the radius r be fixed and analyze the concentration of beliefs on a smaller ball with a radius R_k. The radius R_k needs to be small enough, so we will impose the corresponding upper bound when needed. We start by assuming that $R_k \leq r/2$, thus

$$\mu_k^i(\mathcal{B}_r) \geq 1 - \frac{\int_{\mathcal{B}_r^c} \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\theta)}{\int_{\mathcal{B}_{R_k}} \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\theta)}.$$

Following Proposition 2, it follows that there exists a $\theta \in \mathcal{B}_{R_k}$ such that

$$\mu_k^i(\mathcal{B}_r) \geq 1 - \frac{\int_{\mathcal{B}_r^c} \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\theta)}{\mu_0(\mathcal{B}_{R_k}) \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}}}.$$

Furthermore, we can use the covering of the set in (7) to obtain,

$$\mu_k^i(\mathcal{B}_r) \geq 1 - \frac{\sum_{l=1}^{L-1}\sum_{m=1}^{K_l} \int_{\mathcal{F}_{l,m}} \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}} d\mu_0(\theta)}{\mu_0(\mathcal{B}_{R_k}) \prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}}}$$

$$\geq 1 - \frac{\sum_{l=1}^{L-1}\sum_{m=1}^{K_l} \bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)\mu_0(\mathcal{F}_{l,m})}{\mu_0(\mathcal{B}_{R_k})\bar{g}^i_\theta(\boldsymbol{X}^k)}, \tag{12}$$

where by definition, each $\mathcal{F}_{l,m}$ is contained in a n-Hellinger ball of radius ε_l centered at a point $m \in S_{\varepsilon_l}$

Equation (12) defines a ratio between $\bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)/\bar{g}^i_\theta(\boldsymbol{X}^k)$, where the numerator is defined over the set $\mathcal{F}_{l,m}$ and the denominator with respect to $\theta \in \mathcal{B}_{R_k} \subset \Theta$.

Lemma 3 provides a way to bound term $\bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)/\bar{g}^i_\theta(\boldsymbol{X}^k)$ with high probability. Specifically note that the set $\mathcal{F}_{l,m}$ is inside a ball of radius ε_l centered at $\prod_{j=1}^n P^j_{m,l}$, a covering of the annuli $\mathcal{F}_l$. Moreover, the distribution $\prod_{j=1}^n P^j_\theta$ is inside the ball of radius R_k centered at $\prod_{j=1}^n P^j$. These two balls do not intersect, as $R_k \leq r/2$ and later we will set $\varepsilon_l = r_{l+1}/32$. Thus, the technical hypothesis of Lemma 3 where the distribution $\prod_{j=1}^n P^j_\theta$ is outside the ball $\mathcal{B}_r(\theta_B)$ holds. Note that in this case $r = \varepsilon_l$ and θ_B is the center of one of the covering balls S_{ε_l}.

Thus,

$$\mathbb{P}_\theta\left(\log\frac{\bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)}{\bar{g}^i_\theta(\boldsymbol{X}^k)} \geq y\right) \leq \exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{n}\left(\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)}-\varepsilon_l\right)^2\right), \tag{13}$$

where $P^j_{m,l}$ is the distribution at a point $m \in S_{\varepsilon_l}$, S_{ε_l} is a maximal ε_l separated set of $\mathcal{F}_l$ as in Definition 2. Now, lets analyze the result in (13) with respect to the covering.

$$\begin{aligned}
&\left(\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)}-\varepsilon_l\right)^2 \\
&= \sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta) - 2\varepsilon_l\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)} + \varepsilon_l^2 \\
&\geq \sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta) - 2\varepsilon_l\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)}.
\end{aligned}$$

We have that

$$\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)} \geq \sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j)} - \sqrt{\sum_{j=1}^n h^2(P^j,P^j_\theta)}.$$

Now, lets bound each of the terms above. Initially, note that the true joint distribution of the observations $\prod_{j=1}^n P^j$ is the center of the n-Hellinger ball of radius R_k, and by definition (Proposition 2) $\prod_{j=1}^n P^j_\theta$ is inside the same ball. Therefore, $\sqrt{\sum_{j=1}^n h^2(P^j,P^j_\theta)} \leq R_k$.

On the other hand, recall that the set $\mathcal{F}_{l,m}$ is contained inside a n-Hellinger ball of radius ε_l centered at $m \in S_{\varepsilon_l}$ which is a maximal ε_l-separated subset of the annuli $\mathcal{F}_l = \mathcal{B}_{r_l}/\mathcal{B}_{r_{l+1}}$. Thus, $\prod_{j=1}^n P^j_{m,l} \in \mathcal{F}_l$, and any element of $\mathcal{F}_l$ is at least r_{l+1} away from $\prod_{j=1}^n P^j$, from which it follows that $\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j)} \geq r_{l+1}$, thus

$$\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)} \geq r_{l+1} - R_k \geq \frac{1}{2}r_{l+1},$$

where the last inequality follows by setting $R_k \leq r_l/2$ for $\ell = 1,\cdots,L$. Therefore, so far we have

$$\left(\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)}-\varepsilon_l\right)^2 \geq \frac{1}{4}r^2_{l+1} - 2\varepsilon_l r_l.$$

Next, from setting $\varepsilon_l = r_{l+1}/32$, and $r_l \leq 2r_{l+1}$, and obtain

$$\left(\sqrt{\sum_{j=1}^n h^2(P^j_{m,l},P^j_\theta)}-\varepsilon_l\right)^2 \geq \frac{1}{8}r^2_{l+1},$$

which in turn leads to

$$\mathbb{P}_\theta\left(\log\frac{\bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)}{\bar{g}^i_\theta(\boldsymbol{X}^k)} \geq y\right) \leq \exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{8n}r^2_{l+1}\right). \tag{14}$$

Now, define the set

$$\Gamma^k_{\mathcal{B}_r} = \left\{\boldsymbol{X}^k \middle| \sup_{\substack{l=1,\cdots,L-1\\ m=1,\cdots,K_l}} \log\frac{\bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)}{\bar{g}^i_\theta(\boldsymbol{X}^k)} \geq y\right\},$$

and using the union bound on the coverings defined in Section III-C, we can partition the set $\mathcal{B}^c_r$ in the following way. First, note that $\mathcal{B}_r = \bigcup_{l=1}^{L-1}\mathcal{F}_l$. Recall that $\mathcal{F}_l = \mathcal{B}_{r_l} \setminus \mathcal{B}_{r_{l+1}}$. Moreover, $\mathcal{F}_l = \bigcup_{m\in S}\mathcal{F}_{m,l}$, where each $\mathcal{F}_{m,l} \subseteq \mathcal{B}_{\varepsilon_l}(m)$. Thus, using (14)

$$\begin{aligned}
\mathbb{P}_\theta\left(\Gamma^k_{\mathcal{B}_r}\right) &\leq \sum_{l=1}^{L-1}\sum_{m=1}^{K_l}\mathbb{P}_\theta\left(\log\frac{\bar{g}^i_{\mathcal{F}_{l,m}}(\boldsymbol{X}^k)}{\bar{g}^i_\theta(\boldsymbol{X}^k)} \geq y\right) \\
&\leq \sum_{l=1}^{L-1}\sum_{m=1}^{K_l}\exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{8n}r^2_{l+1}\right) \\
&\leq \sum_{l=1}^{L-1} K_l\exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{8n}r^2_{l+1}\right).
\end{aligned}$$

Now, lets set $y = -\frac{k}{8n}r^2$, and by definition $r \leq r_l$ for all $l = 1,\cdots,L$. Thus

$$\mathbb{P}_\theta\left(\Gamma^k_{\mathcal{B}_r}\right) \leq \sum_{l=1}^{L-1} K_l\exp\left(\frac{4\log n}{1-\delta}-\frac{k}{16n}r^2_{l+1}\right). \tag{15}$$

The probability measure in (15) is computed for $\boldsymbol{X}^k$ distributed according to $\boldsymbol{P}^{\otimes k}_\theta$. Nonetheless, $\boldsymbol{X}^k$ is distributed according to the (slightly different) $\boldsymbol{P}^{\otimes k}$. Our next step is to relate these two measures. First, note that it holds that the total variation distance $D(P^n,Q^n) \geq h^2(P^n,Q^n) = 1-\rho^n(P,Q)$, see for example, [53, Proof of Lemma 1]. Now, it is a fact that the total variation distance between distributions P and Q is upper bounded by $\sqrt{2}h(P,Q)$. Thus, by definition of the total variation distance, for any measurable set B it holds that

$$\sup_B |\mathbb{P}^{\otimes k}_\theta(B) - \mathbb{P}^{\otimes k}(B)|^2 \leq 2h^2(\boldsymbol{P}^{\otimes k}_\theta, \boldsymbol{P}^{\otimes k}).$$

For Hellinger distances [43, Eq. 2.3] it holds

$$h^2(\boldsymbol{P}_\theta^{\otimes k}, \boldsymbol{P}^{\otimes k}) = 1 - \prod_{t=1}^{k}\prod_{j=1}^{n}(1 - h^2(P_\theta^j, P^j)). \tag{16}$$

Next, we are going to use the fact that $(1-(1-x))^n \leq nx$ for $x \in [0,1]$, and $(1-x)^n \geq (1-nx)$. We thus have that

$$1 - \prod_{t=1}^{k}\prod_{j=1}^{n}(1 - h^2(P_\theta^j, P^j)) \leq 1 - k\sum_{j=1}^{n} h^2(P_\theta^j, P^j),$$

and $h^2(\boldsymbol{P}_\theta^{\otimes k}, \boldsymbol{P}^{\otimes k}) \leq k\sum_{j=1}^{n} h^2(P_\theta^j, P^j)$. Then, from the fact that $\theta \in \mathcal{B}_{R_k}$, we have

$$\sup_B \left(\mathbb{P}_\theta(B) - \mathbb{P}^{\otimes k}(B)\right)^2 \leq 2kR_k^2.$$

Setting $R_k \leq \sqrt{\sigma^2/(8k)}$, we obtain $\sup_B \left(\mathbb{P}_\theta^{\otimes k}(B) - \mathbb{P}^{\otimes k}(B)\right)^2 \leq \sigma^2/4$. It is necessary that $R_k \leq \min\left\{\sqrt{\sigma^2/(8k)}, r/2\right\}$. Therefore, we have that

$$\begin{aligned}\mathbb{P}(\Gamma_B^k) &< \mathbb{P}_\theta(\Gamma_B^k) + \sqrt{2kR_k^2} \\ &\leq \sum_{l=1}^{L-1} K_l \exp\left(\frac{4\log n}{1-\delta} - \frac{k}{16n} r_{l+1}^2\right) + \frac{\sigma}{2}.\end{aligned} \tag{17}$$

We are interested in finding a large enough k such that the probability described in (17) is at most σ. Thus, we define

$$N \geq \inf\left\{t \geq 1 \,\middle|\, \sum_{l=1}^{L-1} K_l \exp\left(\frac{4\log n}{1-\delta} - \frac{k}{16n} r_{l+1}^2\right) < \frac{\sigma}{2}\right\}.$$

It follows from (12) that with probability $1-\sigma$ for all $k \geq N$,

$$\begin{aligned}\mu_k^i(\mathcal{B}_r) &\geq 1 - \sum_{l=1}^{L-1}\sum_{m=1}^{K_l} \exp\left(-\frac{k}{8n} r_{l+1}^2\right) \frac{\mu_0(\mathcal{F}_{l,m})}{\mu_0(\mathcal{B}_{R_k})} \\ &\geq 1 - \sum_{l=1}^{L-1} \exp\left(-\frac{k}{8n} r_{l+1}^2\right) \frac{\mu_0(\mathcal{F}_l)}{\mu_0(\mathcal{B}_{R_k})} \\ &\geq 1 - \sum_{l=1}^{L-1} \exp\left(-\frac{k}{8n} r_{l+1}^2\right) \frac{\mu_0(\mathcal{B}_{r_l})}{\mu_0(\mathcal{B}_{R_k})}.\end{aligned}$$

At this point we can use Assumption 3 on the ratio of the initial beliefs on the Hellinger balls $\mathcal{B}_{r_l}$ and $\mathcal{B}_{R_k}$. Then,

$$\begin{aligned}\mu_0(\mathcal{B}_{r_l})/\mu_0(\mathcal{B}_{R_k}) &\leq \exp\left(c_{n,d} r_l / R_k\right) \\ &\leq \exp\left(2c_{n,d} r_{l+1} / \sqrt{\sigma^2/(8k)}\right) \\ &\leq \exp\left(4\sqrt{2} c_{n,d} \sqrt{k} r_{l+1}/\sigma\right).\end{aligned}$$

Finally, we can conclude that

$$\mu_k^i(\mathcal{B}_r) \geq 1 - \sum_{l=1}^{L-1} \exp\left(-\frac{k}{8n} r_{l+1}^2 + 4\sqrt{2} c_{n,d} \frac{\sqrt{k} r_{l+1}}{\sigma}\right).$$

Moreover, for any $k \geq \left(64\sqrt{2} c_{n,d} n/(\sigma r_{l+1})\right)^2$, it holds that

$$\mu_k^i(\mathcal{B}_r) \geq 1 - \sum_{l=1}^{L-1} \exp\left(-\frac{k}{16n} r_{l+1}^2\right).$$

■

Analogous to Theorem 5 in Part I [52], Theorem 4 provides a probabilistic concentration result for the agents' beliefs around a Hellinger ball of radius r with center at θ^* for sufficiently large k. We provide an explicit number of iterations after which an exponential concentration occurs. Moreover, the rate at which this happens is proportional to the radius r of a ball around the optimal hypotheses.

IV. Experimental Results

This section shows a number of experimental results for the distributed estimation of network-wide parameters for various network topologies and observational models. In particular, we focus on observational models from the exponential family of distributions.

Table I recalls the results from Part I [52] for a number of distributed estimation problems with likelihood models coming from exponential families.

Particularly, we describe the relation between the distribution of the observations, the parameter space and the belief distributions. Moreover, we provide explicit relations between the parameters in the canonical form and the corresponding parameters of the beliefs as in Proposition 2 in Part I [52], where it was shown that the belief update rule (4) can be written as

$$\chi_{k+1}^i = \sum_{j=1}^{n} a_{ij}\chi_k^j + T^i(x_{k+1}^i).$$

when the set of probability distributions whose density can be represented as $p_\theta(x) = H(x)\exp(M(\theta)'T(x))$, for specific functions $H(\cdot)$, $M(\cdot)$ and $T(\cdot)$ where $M(\theta) = [M(\theta^1), M(\theta^2), \ldots, M(\theta^s)]'$ depends on the density parameters and $T(\cdot)$ depends on the observations.

We explore four different distributed network-wide estimation problems:

- Figure 2: Estimating the variance with Gaussian observations with local knowledge of private means.
- Figure 3: Estimating the mean and variance with Gaussian observations without knowledge of local means or variances.
- Figure 4: Estimating the mean with heterogeneous Bernoulli observations.

Simulations for estimating the mean with Gaussian observations with local knowledge of private variances, the case of estimating the mean with heterogeneous Bernoulli observations, and estimating the mean with heterogeneous Poisson observations, are shown in the appendix.

For each of the figures described above, we measure the performance of the proposed algorithm using its normalized distance to optimality and the distance to consensus, defined as follows

$$\text{Optimality: } \frac{|F(\theta_k) - F(\theta^*)|}{|F(\theta_0) - F(\theta^*)|}, \text{Consensus: } \|\mathcal{L}\theta_k\|_2^2,$$

where $\theta_k = (\theta_k^1, \theta_k^2, \cdots, \theta_k^n)$ is the aggregation of all the current parameters estimation for each of the agents, the function $F(\theta_k)$ is defined as $F(\theta_k) = \sum_{i=1}^{n} D_{KL}(P^i \| P_{\theta_k^i}^i)$, and $\mathcal{L}$ is the graph Laplacian of the communication graph. We have

Observations X_k^i	Parameter Space Θ	Beliefs Distribution	$T(x)$	$M(\theta)$	Belief Parameters
Bern(θ^i)	$\{\theta \in [0,1]\}$	Beta(α_k^i, β_k^i)	x	$\log \frac{\theta}{1-\theta}$	$\begin{bmatrix} \alpha_k^i = \chi_k^i + 1 \\ \beta_k^i = \chi_k^i + \nu_k^i + 1 \end{bmatrix}$
Binomial(θ^i, m^i)	$\{\theta \in [0,1]\}$	Beta(α_k^i, β_k^i)	x	$\log \frac{\theta}{1-\theta}$	$\begin{bmatrix} \alpha_k^i = \chi_k^i + 1 \\ \beta_k^i = \chi_k^i + m^i\nu_k^i + 1 \end{bmatrix}$
Multinomial(θ^i, m^i)	$\{\theta \in [0,1]^d, \sum \theta_i = 1\}$	Dirichlet($\alpha_k^i \in \mathbb{R}_+^d$)	x	$\log \theta$	$\alpha_k^i = \chi_k^i + 1$
Poisson(θ^i)	$\{\theta > 0\}$	Gamma(α_k^i, β_k^i)	x	$\log \theta$	$\begin{bmatrix} \alpha_k^i = \chi_k^i + 1 \\ \beta_k^i = \nu_k^i \end{bmatrix}$
Exp(θ^i)	$\{\theta > 0\}$	Gamma(α_k^i, β_k^i)	x	$-\theta$	$\begin{bmatrix} \alpha_k^i = \nu_k^i + 1 \\ \beta_k^i = \chi_k^i \end{bmatrix}$
$\mathcal{N}(\theta^i \vert (\sigma^i)^2)$	$\{\theta \in \mathbb{R}\}$	$\mathcal{N}(\bar{\theta}_k^i, (\bar{\sigma}_k^i)^2)$	x	$\frac{\theta}{(\sigma^i)^2}$	$\begin{bmatrix} \bar{\theta}_k^i = \frac{\chi_k^i}{\nu_k^i} \\ (\bar{\sigma}_k^i)^2 = \frac{(\sigma^i)^2}{\nu_k^i} \end{bmatrix}$
$\mathcal{N}(\tau^i \vert \theta^i)$	$\{\tau > 0\}$	Gamma(α_k^i, β_k^i)	$\frac{1}{2}(x-\theta^i)^2$	$-\tau$	$\begin{bmatrix} \alpha_k^i = \frac{\nu_k^i}{2} + 1 \\ \beta_k^i = \chi_k^i \end{bmatrix}$
$\mathcal{N}(\theta^i, \tau^i)$	$\{\theta \in \mathbb{R}, \tau > 0\}$	$\mathcal{N}$-Gamma($\bar{\theta}_k^i, \bar{\tau}_k^i, \alpha_k^i, \beta_k^i$)	$\begin{bmatrix} x^2 \\ x \\ \frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -\frac{1}{2}\tau \\ \tau\theta \\ \log \tau \end{bmatrix}$	$\begin{bmatrix} \alpha_k^i = [\chi_k^i]_1 - \frac{1}{2} \\ \beta_k^i = \frac{1}{2}[\chi_k^i]_2 - \frac{1}{2}\frac{([\chi_k^i]_3)^2}{\nu_k^i} \\ \theta_k^i = \frac{[\chi_k^i]_3}{\nu_k^i} \\ \lambda_k^i = [\chi_k^i]_4 \end{bmatrix}$

TABLE I: Parameter Descriptions for Distributed Learning on the Exponential Family.

used the graph Laplacian as a measure to distance of consensus since by definition the set where $\theta_k^1 = \theta_k^2 = \cdots = \theta_k^n$, i.e. consensus, is null space of the matrix $\mathcal{L}$.

Finally, we present the results for five classes of networks, namely: complete graphs, cycle graphs, path graphs, star graphs, and Erdős-Rényi random graphs. For each network class, we show the performance for 10 agents, 100 agents, and 1000 agents.

In all experimental results, the predicted geometric convergence rate is observed. Moreover, as the number of agents in the network increases, the effects of the network topology become more evident. Mainly, for highly connected graphs such as the complete graph or the Erdős-Rényi, the distance to optimality and consensus decays faster. One interesting observation is that contrary to what was expected, the performance of the proposed algorithm on graphs with a star topology is worst in most cases. This can be explained by the fact that given that the agents are in a well-connected graph, they are oblivious to the topology of the network and thus cannot exploit the network structure. The central node does not know it is a central node, and similarly for the other agents.

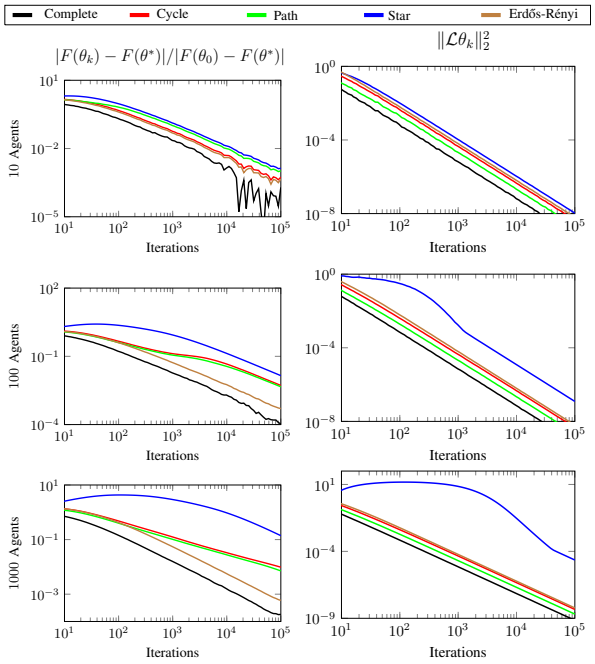


Fig. 2: Optimality and distance to consensus for the distributed estimation of a network-wide **variance** parameter, from Gaussian observations, for various graph topologies (complete, cycle, path, star and Erdős-Rényi) of increasing size (10 agents, 100 agents, and 1000 agents).

V. Conclusions

We have proposed an algorithm for distributed learning with compact hypothesis sets. Our algorithm may be viewed as a distributed version of Stochastic Mirror Descent applied to the problem of minimizing the sum of Kullback-Leibler divergences. Our results show non-asymptotic geometric convergence rates for the beliefs concentration around the true hypothesis.

Future work should explore how variations on stochastic approximation algorithms will produce new non-Bayesian update rules for more general problems. Furthermore, we have modeled interactions between agents as exchanges of local probability distributions (i.e., beliefs) between neighboring nodes in a graph. It remains open to understanding to what extent this can be reduced when agents transmit only an approximate summary of their beliefs. We anticipate that future work will also consider the effect of parametric approximations allowing nodes to communicate only a finite number of parameters coming from Gaussian Mixture Models or Particle Filters.

Acknowledgment

We would like to acknowledge support for this project from the National Science Foundation under grant no. CPS 15-44953 and by the Office of Naval Research under grant no. N00014-17-1-2195.

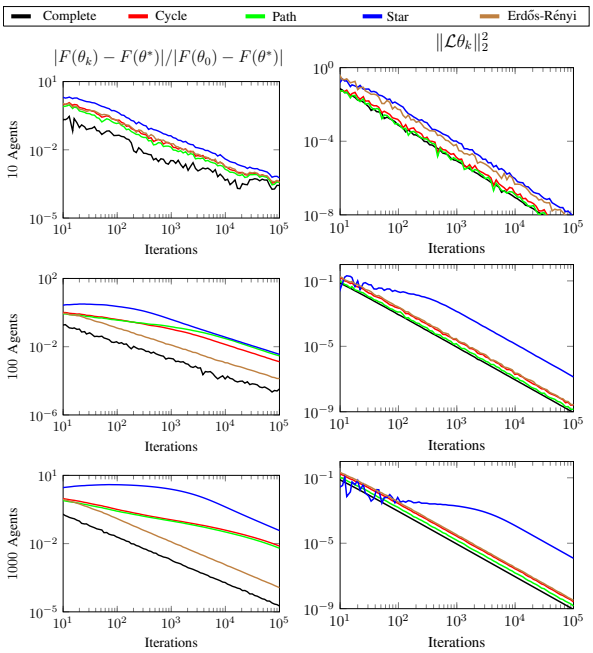


Fig. 3: Optimality and distance to consensus for the distributed estimation of a network-wide **mean and variance** parameters, from Gaussian observations, for various graph topologies (complete, cycle, path, star and Erdős-Rényi) of increasing size (10 agents, 100 agents and 1000 agents).

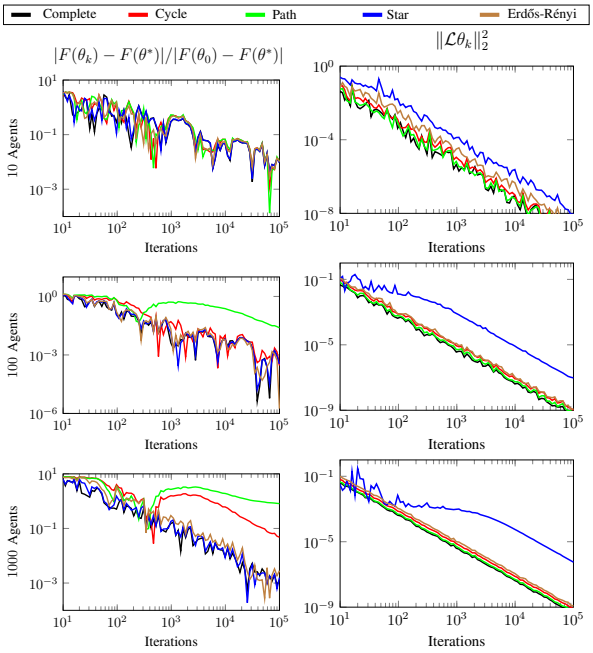


Fig. 4: Optimality and distance to consensus for the distributed estimation of a network-wide parameter of Bernoulli observations for various graph topologies (complete, cycle, path, star and Erdős-Rényi) of increasing size (10 agents, 100 agents, and 1000 agents).

APPENDIX

Network Topology	Mixing Time
Cycle [56, Section 5.3.1]	$O(n^2)$
Path [57], [58]	$O(n^2)$
Star Graph [58]	$O(1)$
Dumbbell Graph [59]	$O(n^2)$
Lollipop [60]	$O(n^2)$
Complete Binary Tree [56, Section 5.3.4]	$O(n)$
k-d Hypercube $\{0,1\}^k$ [56, Section 5.3.3]	$O(k \log k + k)$
L-Lattice on $\mathcal{Z}_n \times \mathcal{Z}_n$ [61], [62]	$O(n^2)$
k-d Grid [61], [62]	$O(2k^2 n^{2/k} \log n)$
k-d Torus [56, Section 5.3.3]	$O(k^2 n^2)$
Eulerian Graph [63]	$O(\|E\|^2)$
Lazy Eulerian with degree d-degree [64]	$O(n\|E\|)$
Eulerian: d-degree, max-degree weights and expansion [63]	$O(n^2 d)$
Geometric Random Graph: $\mathcal{G}^d(n,r)$ [65]	$O(r^{-2} \log n)$
Geometric Random Graph: $\mathcal{G}^2(n, \Omega(\text{polylog}(n)))$ [66]	$O(\text{polylog}(n))$
Erdős-Rényi: $\mathcal{G}(n, c/n)$, $c > 1$ [67]	$O(\log^2 n)$
Erdős-Rényi: $\mathcal{G}(n, (1+\delta)/n)$, $\delta^3 n \to \infty$ [68]	$O((1/\delta^3)\log^2(\delta^3 n))$
Erdős-Rényi: $\mathcal{G}(n, 1/n)$ [69]	$O(n)$
Newman-Watts (small-world) Graph [70]	$O(\log^2 n)$
Expander Graph [71]	$O(\log n)$
Any Connected Undirected Graph with Metropolis weights [72]	$O(n^2)$

TABLE II: Upper bounds on the mixing time for various graph topologies. The mixing times, in turn, serve as bounds for the spectral gap of the mixing matrices.

PROOF OF LEMMA 3

Proof: By the Markov inequality, it follows that

$$\mathbb{P}_\theta\left[\log \frac{\bar{g}_B^i(\boldsymbol{X}^k)}{\bar{g}_\theta^i(\boldsymbol{X}^k)} \geq y\right] \leq \exp(-y/2)\mathbb{E}_\theta\left[\sqrt{\frac{\bar{g}_B^i(\boldsymbol{X}^k)}{\bar{g}_\theta^i(\boldsymbol{X}^k)}}\right]$$
$$= \exp(-y/2)\int_{\boldsymbol{\mathcal{X}}^k} \sqrt{\frac{\bar{g}_B^i(\boldsymbol{x}^k)}{\bar{g}_\theta^i(\boldsymbol{x}^k)}} g_\theta(\boldsymbol{x}^k) d\lambda^{\otimes kn}(\boldsymbol{x}^k).$$

Initially, we will use the multivariate Jensen's inequality[3] [73], with $x^{[A^{k-t}]_{ij}}$ being a concave function and $1/\mu_0(B)\int_B d\mu_0 = 1$. The function $p_\theta^j(x_t^j)$ maps the parameter θ to a vector with kn entries, which are then mapped into the geometric mean via the function $\prod_{t=1}^{k}\prod_{j=1}^{n}(\cdot)^{[A^{k-t}]_{ij}}$. Such a relation is also a consequence of the generalized Hölder's inequality [74] with the assumption that the matrix A is doubly stochastic. Thus, we have that

$$\bar{g}_B^i(\boldsymbol{x}^k) = \frac{1}{\mu_0(B)}\int_B \prod_{t=1}^{k}\prod_{j=1}^{n}\left(p_\theta^j(x_t^j)\right)^{[A^{k-t}]_{ij}} d\mu_0(\theta)$$
$$\leq \prod_{t=1}^{k}\prod_{j=1}^{n}\left(\frac{1}{\mu_0(B)}\int_B p_{\hat{\theta}}^j(x_t^j) d\mu_0(\hat{\theta})\right)^{[A^{k-t}]_{ij}}.$$

Therefore,

$$\sqrt{\frac{\bar{g}_B^i(\boldsymbol{x}^k)}{\bar{g}_\theta^i(\boldsymbol{x}^k)}} \leq \sqrt{\frac{\prod_{t=1}^{k}\prod_{j=1}^{n}\left(\frac{1}{\mu_0(B)}\int_B p_{\hat{\theta}}^j(x) d\mu_0(\hat{\theta})\right)^{[A^{k-t}]_{ij}}}{\prod_{t=1}^{k}\prod_{j=1}^{n} p_\theta^j(x_t^j)^{[A^{k-t}]_{ij}}}}$$

[3]For a concave function ϕ and $\int_\Omega f(x)dx = 1$, it holds that $\int_\Omega \phi(g(x))f(x)dx \leq \phi\left(\int_\Omega g(x)f(x)dx\right)$. The function f is the density of $d\mu_0/\mu_0(B)$, the function ϕ is the geometric average $\prod_{t=1}^{k}\prod_{j=1}^{n}(\cdot)^{[A^{k-t}]_{ij}}$, and the function g maps θ to the vector with entries $p_\theta^j(x_t^j)$.

$$= \sqrt{\prod_{t=1}^{k}\prod_{j=1}^{n}\left(\frac{\frac{1}{\mu_0(B)}\int_B p_{\hat{\theta}}^j(x_t^j)d\mu_0(\hat{\theta})}{p_\theta^j(x_t^j)}\right)^{[A^{k-t}]_{ij}}}$$

Next, applying the same argument with the multivariate Jensen's inequality and $x^{[A^{k-t}]_{ij}}$ but now with respect to the measure $g_\theta(\boldsymbol{x}^k)d\lambda^{\otimes kn}(\boldsymbol{x}^k)$ we obtain

$$\begin{aligned}
&\mathbb{P}_\theta\left[\log\frac{\bar{g}_B^i(\boldsymbol{X}^k)}{\bar{g}_\theta^i(\boldsymbol{X}^k)}\geq y\right] \leq \exp(-y/2)\times \\
&\quad\times\int_{\boldsymbol{\mathcal{X}}^k}\sqrt{\prod_{t=1}^{k}\prod_{j=1}^{n}\left(\frac{\frac{1}{\mu_0(B)}\int_B p_{\hat{\theta}}^j(x_t^j)d\mu_0(\hat{\theta})}{p_\theta^j(x_t^j)}\right)^{[A^{k-t}]_{ij}}} \\
&\times g_\theta(\boldsymbol{x}^k)d\lambda^{\otimes kn}(\boldsymbol{x}^k) \\
&\leq \exp(-y/2)\times \\
&\times\prod_{t=1}^{k}\prod_{j=1}^{n}\left(\int_{\boldsymbol{\mathcal{X}}}\sqrt{\frac{\frac{1}{\mu_0(B)}\int_B p_\theta^j(x_t^j)d\mu_0(\theta)}{p_\theta^j(x_t^j)}}p_\theta^j(x_t^j)d\lambda^{\otimes n}(\boldsymbol{x})\right)^{[A^{k-t}]_{ij}}.
\end{aligned}$$

Thus, we obtain

$$\mathbb{P}_\theta\left[\log\frac{\bar{g}_B^i(\boldsymbol{X}^k)}{\bar{g}_\theta^i(\boldsymbol{X}^k)}\geq y\right] \leq \exp\left(-\frac{y}{2}\right)\prod_{t=1}^{k}\prod_{j=1}^{n}\rho\left(P_B^j,P_\theta^j\right)^{[A^{k-t}]_{ij}},$$

where P_B^j is the measure with Radon-Nikodym derivative $g_B^j(x)=\frac{1}{\mu_0(B)}\int_B p_\theta^j(x)d\mu_0(\theta)$ with respect to λ. Now, recall that the Hellinger distance is bounded by above by 1. Thus,

$$\prod_{t=1}^{k}\prod_{j=1}^{n}\rho(P_B^j,P_\theta^j)^{[A^{k-t}]_{ij}} \leq \exp\left(-\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij}h^2(P_B^j,P_\theta^j)\right).$$

Moreover we can add and subtract $\sum_{t=1}^{k}\sum_{j=1}^{n}\frac{1}{n}h^2(P_B^j,P_\theta^j)$, thus,

$$\begin{aligned}
&\prod_{t=1}^{k}\prod_{j=1}^{n}\rho(P_B^j,P_\theta^j)^{[A^{k-t}]_{ij}} \\
&\leq \exp\left(-\sum_{t=1}^{k}\sum_{j=1}^{n}\left([A^{k-t}]_{ij}-\frac{1}{n}\right)h^2(P_B^j,P_\theta^j)\right. \\
&\left.-\frac{k}{n}\sum_{j=1}^{n}h^2(P_B^j,P_\theta^j)\right).
\end{aligned}$$

Additionally, from Lemma 1,

$$\sum_{t=1}^{k}\sum_{j=1}^{n}\left([A^{k-t}]_{ij}-\frac{1}{n}\right)h^2(P_B^j,P_\theta^j)\leq\frac{4\log n}{1-\delta},$$

from which we can conclude that

$$\mathbb{P}_\theta\left[\log\frac{\bar{g}_B^i(\boldsymbol{X}^k)}{\bar{g}_\theta^i(\boldsymbol{X}^k)}\geq y\right] \leq \exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{n}\sum_{j=1}^{n}h^2(P_B^j,P_\theta^j)\right).$$

Note that the term $\sum_{j=1}^{n}h^2(P_B^j,P_\theta^j)$ above, is the squared distance (c.f. Definition 1) between the distribution defined on the set $B\subset\mathcal{B}_r(\theta_B)$ and $\theta\in\Theta/\mathcal{B}_r(\theta_B)$. Analogously to [43, Propositon 5], it follows that:

$$\sqrt{\sum_{j=1}^{n}h^2(P_B^j,P_\theta^j)}\geq\sqrt{\sum_{j=1}^{n}h^2\left(P_{\theta_B}^j,P_\theta^j\right)}-r,$$

thus

$$\begin{aligned}
&\mathbb{P}_\theta\left[\log\frac{\bar{g}_B(\boldsymbol{X}^k)}{\bar{g}_\theta(\boldsymbol{X}^k)}\geq y\right] \\
&\leq\exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{n}\left(\sqrt{\sum_{j=1}^{n}h^2\left(P_{\theta_B}^j,P_\theta^j\right)}-r\right)^2\right).
\end{aligned}$$

■

Additional Numerical Results

Next, we present two additional numerical results:

- Figure 5: Estimating the mean with heterogeneous Poisson observations.
- Figure 6: Estimating the mean with heterogeneous Exponential observations.

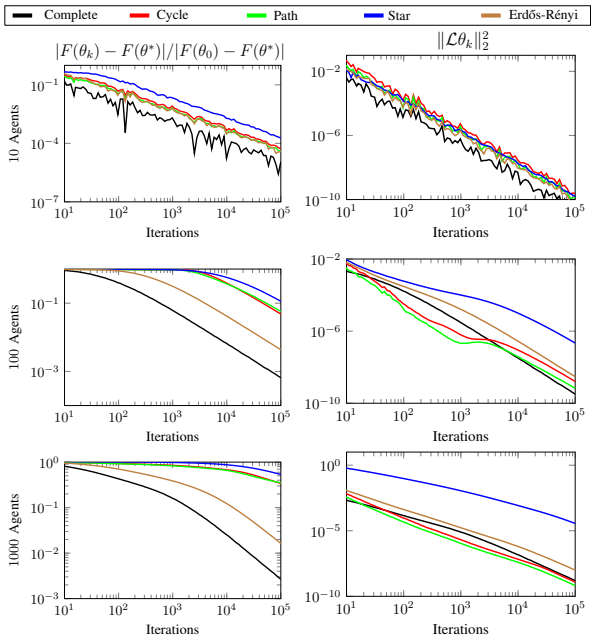


Fig. 5: Optimality and distance to consensus for the distributed estimation of a network-wide parameter of Poisson observations for various graph topologies (complete, cycle, path, star and Erdős-Rényi) of increasing size (10 agents, 100 agents, and 1000 agents).

References

[1] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[2] K. Rahnama Rad and A. Tahbaz-Salehi, "Distributed parameter estimation in networks," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 5050–5055, 2010.

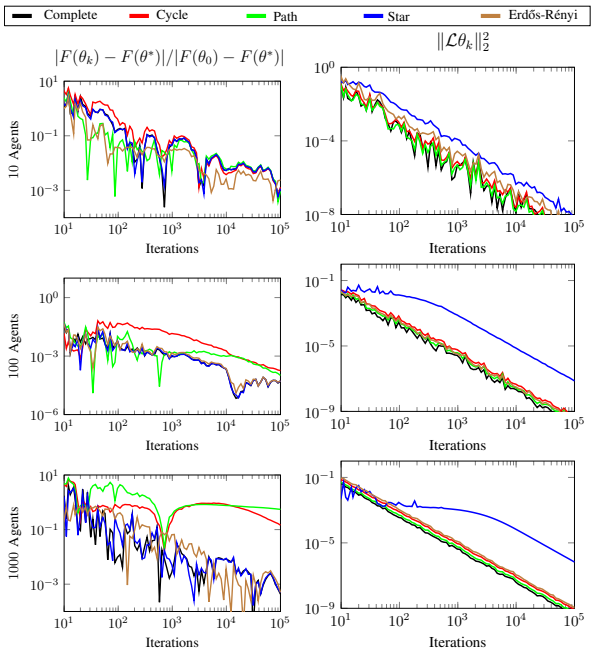


Fig. 6: Optimality and distance to consensus for the distributed estimation of a network-wide parameter of Exponential observations for various graph topologies (complete, cycle, path, star and Erdős-Rényi) of increasing size (10 agents, 100 agents, and 1000 agents).

[3] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron, "Distributed bayesian hypothesis testing in sensor networks," in *Proceedings of the American Control Conference*, pp. 5369–5374, 2004.

[4] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Networked Embedded Sensing and Control*, pp. 169–182, Springer, 2006.

[5] R. J. Aumann, "Agreeing to disagree," *The Annals of Statistics*, vol. 4, no. 6, pp. 1236–1239, 1976.

[6] V. Borkar and P. P. Varaiya, "Asymptotic agreement in distributed estimation," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 650–655, 1982.

[7] J. N. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Transactions on Automatic Control*, vol. 29, no. 1, pp. 42–50, 1984.

[8] C. Genest, J. V. Zidek, *et al.*, "Combining probability distributions: A critique and an annotated bibliography," *Statistical Science*, vol. 1, no. 1, pp. 114–135, 1986.

[9] R. Cooke, "Statistics in expert resolution: A theory of weights for combining expert opinion," in *Statistics in Science* (R. Cooke and D. Costantini, eds.), vol. 122 of *Boston Studies in the Philosophy of Science*, pp. 41–72, Springer Netherlands, 1990.

[10] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[11] G. L. Gilardoni and M. K. Clayton, "On reaching a consensus using degroot's iterative pooling," *The Annals of Statistics*, vol. 21, no. 1, pp. 391–401, 1993.

[12] J. A. Gubner, "Distributed estimation and quantization," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1456–1459, 1993.

[13] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1631–1639, 2005.

[14] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[15] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.

[16] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.

[17] E. Mossel and O. Tamuz, "Efficient bayesian learning in social networks with gaussian estimators," *arXiv preprint arXiv:1002.0747*, 2010.

[18] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[19] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Business School Research Paper*, no. 13-28, 2013.

[20] S. Shahrampour and A. Jadbabaie, "Exponentially fast parameter estimation in networks using distributed dual averaging," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 6196–6201, 2013.

[21] B. Golub and M. O. Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, pp. 112–149, 2010.

[22] D. Acemoglu, A. Nedić, and A. Ozdaglar, "Convergence of rule-of-thumb learning rules in social networks," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 1714–1720, 2008.

[23] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[24] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[25] A. Olshevsky, "Linear time average consensus on fixed graphs and implications for decentralized optimization and multi-agent control," *preprint arXiv:1411.4186*, 2014.

[26] E. Mossel, A. Sly, and O. Tamuz, "Asymptotic learning on bayesian social networks," *Probability Theory and Related Fields*, vol. 158, no. 1-2, pp. 127–157, 2014.

[27] A. Lalitha, T. Javidi, and A. D. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6161–6179, 2018.

[28] L. Qipeng, F. Aili, W. Lin, and W. Xiaofan, "Non-bayesian learning in social networks with time-varying weights," in *30th Chinese Control Conference (CCC)*, pp. 4768–4771, 2011.

[29] L. Qipeng, Z. Jiuhua, and W. Xiaofan, "Distributed detection via bayesian updates and consensus," in *34th Chinese Control Conference (CCC)*, pp. 6992–6997, 2015.

[30] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," *IEEE Transactions on Automatic Control*, vol. 61, pp. 3256–3268, Nov 2016.

[31] S. Shahrampour, M. Rahimian, and A. Jadbabaie, "Switching to learn," in *Proceedings of the American Control Conference*, pp. 2918–2923, 2015.

[32] M. A. Rahimian, S. Shahrampour, and A. Jadbabaie, "Learning without recall by random walks on directed graphs," *preprint arXiv:1509.04332*, 2015.

[33] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed non-bayesian learning," *preprint arXiv:1508.05161*, 2015.

[34] A. Nedić, A. Olshevsky, and C. A. Uribe, "Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs," in *Proceedings of the American Control Conference*, pp. 5884–5889, 2015.

[35] A. Nedić, A. Olshevsky, and C. A. Uribe, "Network independent rates in distributed learning," in *Proceedings of the American Control Conference*, pp. 1072–1077, 2016.

[36] L. Su and N. H. Vaidya, "Asynchronous distributed hypothesis testing in the presence of crash failures," *University of Illinois at Urbana-Champaign, Tech. Rep*, 2016.

[37] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, "A theory of non-Bayesian social learning," *Econometrica*, vol. 86, no. 2, pp. 445–490, 2018.

[38] A. Mitra, J. A. Richards, and S. Sundaram, "A new approach for distributed hypothesis testing with extensions to Byzantine-resilience," in *American Control Conference (ACC)*, pp. 261–266, IEEE, 2019.

[39] A. Mitra, J. A. Richards, and S. Sundaram, "A communication-efficient algorithm for exponentially fast non-Bayesian learning in networks," in *IEEE 58th Conference on Decision and Control (CDC)*, pp. 8347–8352, IEEE, 2019.

[40] J. Z. Hare, C. A. Uribe, L. Kaplan, and A. Jadbabaie, "Non-bayesian social learning with uncertain models," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4178–4193, 2020.

[41] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Distributed detection and estimation in wireless sensor networks," *preprint arXiv:1307.1448*, 2013.

[42] A. Nedić, A. Olshevsky, and C. A. Uribe, "A tutorial on distributed (non-bayesian) learning: Problem, algorithms and results," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6795–6801, Dec 2016.

[43] L. Birgé, "About the non-asymptotic behaviour of bayes estimators," *Journal of Statistical Planning and Inference*, vol. 166, pp. 67–77, 2015.

[44] S. Ghosal, "A review of consistency and convergence of posterior distribution," in *Varanashi Symposium in Bayesian Inference, Banaras Hindu University*, 1997.

[45] L. Schwartz, "On bayes procedures," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 4, no. 1, pp. 10–26, 1965.

[46] S. Ghosal, J. K. Ghosh, and A. W. Van Der Vaart, "Convergence rates of posterior distributions," *Annals of Statistics*, pp. 500–531, 2000.

[47] S. Ghosal, A. Van Der Vaart, *et al.*, "Convergence rates of posterior distributions for noniid observations," *The Annals of Statistics*, vol. 35, no. 1, pp. 192–223, 2007.

[48] V. Rivoirard, J. Rousseau, *et al.*, "Posterior concentration rates for infinite dimensional exponential families," *Bayesian Analysis*, vol. 7, no. 2, pp. 311–334, 2012.

[49] C. A. Uribe, A. Olshevsky, and A. Nedić, "Non-asymptotic concentration rates in cooperative learning part i: Variational non-bayesian social learning," *Submitted*, 2020.

[50] M. Rabbat, "Multi-agent mirror descent for decentralized stochastic optimization," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*, pp. 517–520, IEEE, 2015.

[51] J. Li, G. Li, Z. Wu, and C. Wu, "Stochastic mirror descent method for distributed multi-agent optimization," *Optimization Letters*, pp. 1–19, 2016.

[52] C. A. Uribe, A. Olshevsky, and A. Nedić, "Non-asymptotic concentration rates in cooperative learning part ii: Inference on compact hypothesis sets," *Submitted*, 2020.

[53] L. LeCam, "Convergence of estimates under dimensionality restrictions," *The Annals of Statistics*, pp. 38–53, 1973.

[54] A. W. Van der Vaart, *Asymptotic statistics*, vol. 3. Cambridge university press, 2000.

[55] M. H. Protter, B. Charles Jr, *et al.*, *A first course in real analysis*. Springer Science & Business Media, 2012.

[56] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, Providence, 2009.

[57] S. Ikeda, I. Kubo, and M. Yamashita, "The hitting and cover times of random walks on finite graphs using local degree information," *Theoretical Computer Science*, vol. 410, no. 1, pp. 94–100, 2009.

[58] A. Beveridge and M. Wang, "Exact mixing times for random walks on trees," *Graphs and Combinatorics*, vol. 29, no. 4, pp. 757–772, 2013.

[59] R. Kannan, L. Lovász, and R. Montenegro, "Blocking conductance and mixing in random walks," *Combinatorics, Probability and Computing*, vol. 15, no. 4, pp. 541–570, 2006.

[60] D. Aldous and J. Fill, "Reversible Markov chains and random walks on graphs," 2002.

[61] C. Avin and G. Ercal, "Bounds on the mixing time and partial cover of ad-hoc and sensor networks.," in *EWSN*, pp. 1–12, 2005.

[62] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, "The electrical resistance of a graph captures its commute and cover times," *Computational Complexity*, vol. 6, no. 4, pp. 312–340, 1996.

[63] R. Montenegro, "The simple random walk and max-degree walk on a directed graph," *Random Structures & Algorithms*, vol. 34, no. 3, pp. 395–407, 2009.

[64] L. Boczkowski, Y. Peres, and P. Sousi, "Sensitivity of mixing times in Eulerian digraphs," *arXiv preprint arXiv:1603.05639*, 2016.

[65] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Mixing times for random walks on geometric random graphs.," in *ALENEX/ANALCO*, pp. 240–249, 2005.

[66] C. Avin and G. Ercal, "On the cover time and mixing time of random geometric graphs," *Theoretical Computer Science*, vol. 380, no. 1-2, pp. 2–22, 2007.

[67] I. Benjamini, G. Kozma, and N. Wormald, "The mixing time of the giant component of a random graph," *Random Structures & Algorithms*, vol. 45, no. 3, pp. 383–407, 2014.

[68] J. Ding, E. Lubetzky, Y. Peres, *et al.*, "Mixing time of near-critical random graphs," *The Annals of Probability*, vol. 40, no. 3, pp. 979–1008, 2012.

[69] A. Nachmias and Y. Peres, "Critical random graphs: diameter and mixing time," *The Annals of Probability*, pp. 1267–1286, 2008.

[70] L. Addario-Berry and T. Lei, "The mixing time of the Newman-Watts small-world model," *Advances in Applied Probability*, vol. 47, no. 1, pp. 37–56, 2015.

[71] R. Durrett, *Random Graph Dynamics*. Cambridge University Press, UK, 2007.

[72] A. Olshevsky, "Linear time average consensus and distributed optimization on fixed graphs," *SIAM Journal on Control and Optimization*, vol. 55, no. 6, pp. 3990–4014, 2017.

[73] M. D. Perlman, "Jensen's inequality for a convex vector-valued function on an infinite-dimensional space," *Journal of Multivariate Analysis*, vol. 4, no. 1, pp. 52–65, 1974.

[74] W. H. Yang, "On generalized hoder inequality," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 16, no. 5, pp. 489–498, 1991.

César A. Uribe received the M.Sc. degrees in systems and control from the Delft University of Technology, Delft, The Netherlands, and in applied mathematics from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2013 and 2016, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, in 2018.

He is currently the Louis Owen Jr. Chair and Assistant Professor with the Department of Electrical and Computer Engineering Department at Rice University, Houston, TX, USA. From 2018 to 2020, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include distributed learning and optimization, decentralized control, algorithm analysis, and computational optimal transport

Alex Olshevsky received the B.S. degrees in applied mathematics and electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, both in 2004, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2006 and 2010, respectively.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA. His research interests include control systems, optimization, and network science. Dr. Olshevsky received the National Science Foundation CAREER Award, the Air Force Young Investigator Award, the ICS Prize from INFORMS for best paper on the interface of operations research and computer science, and the SIAM paper prize for annual paper from the SIAM Journal on Control and Optimization chosen to be reprinted in SIAM Review.

Angelia Nedić (Member, IEEE) received the Ph.D. degree in computational mathematics and mathematical physics from Moscow State University, Moscow, Russia, in 1994, and the Ph.D. degree in electrical and computer science engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002.

She has worked as a Senior Engineer with BAE Systems North America, Arlington, VA, USA, and the Advanced Information Technology Division, Burlington, MA, USA. She has been a Willard Scholar Faculty Member with the University of Illinois at Urbana–Champaign, Champaign, IL, USA.

She is currently a Faculty Member with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA. Her general research interests include optimization, large-scale complex systems dynamics, variational inequalities, and games. Dr. Nedić was a recipient (jointly with her coauthors) of the best paper awards at the Winter Simulation Conference 2013 and the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) 2015.

Non-asymptotic Concentration Rates in Cooperative Learning Part I: Variational Non-Bayesian Social Learning

César A. Uribe, Alex Olshevsky, and Angelia Nedić

***Abstract*— We study the problem of cooperative inference where a group of agents interacts over a network and seeks to estimate a joint parameter that best explains a set of network-wide observations using local information only. Agents do not know the network topology or the observations of other agents. We explore a variational interpretation of the Bayesian posterior and its relation to the stochastic mirror descent algorithm to prove that, under appropriate assumptions, the beliefs generated by the proposed algorithm concentrate around the true parameter exponentially fast. In Part I of this two-part paper series, we focus on providing a variational approach to distributed Bayesian filtering. Moreover, we develop computationally efficient algorithms for observation models in exponential families. We provide a novel non-asymptotic belief concentration analysis for distributed non-Bayesian learning on finite hypothesis sets. This new analysis is the basis for the results presented in Part II. We provide the first non-asymptotic belief concentration rate analysis for distributed non-Bayesian learning over networks on compact hypothesis sets in Part II. Additionally, we provide extensive numerical analysis for various distributed inference tasks on networks for observational models in the exponential distribution families.**

***Index Terms*— Distributed inference, non-Bayesian social learning, estimation over networks, non-asymptotic rates.**

I. Introduction

The increasing amount of data generated by recent applications of distributed systems such as social media, sensor networks, and cloud-based databases has brought considerable attention to distributed data processing, in particular the design of distributed algorithms that take into account the communication constraints and make collective decisions in a distributed manner [1]–[11]. In a distributed system, interactions between agents are usually constrained by the network structure, and agents can only use locally available information. This contrasts with centralized approaches where all information and computation resources are available at a single location [12]–[15].

One traditional problem in decision-making is that of parameter estimation. Given a set of noisy observations coming from a joint distribution, one would like to estimate a parameter or distribution that minimizes a specific loss function. For example, Maximum a Posteriori (MAP) or Minimum Least Squared Error (MLSE) estimators fit a parameter to some model of the observations. Both MAP and MLSE estimators require some form of Bayesian posterior computation based on models that explain the observations for a given parameter. Computation of such *a posteriori* distribution depends on having exact models about the likelihood of the corresponding observations. This is one of the main difficulties of using Bayesian approaches in a distributed setting. Different agents might have observations from different distributions. If sharing data were possible, sharing local models would be necessary for information aggregation and inference. Moreover, the agents are assumed oblivious to the network topology. Thus doubly counting and data indexing could be challenging in large networks. A fully Bayesian approach is not possible since full knowledge of the network structure or other agents' likelihood models may not be available [16]–[18].

Following the seminal work of Jadbabaie et al. in [1], [19], [20], there have been many studies of distributed non-Bayesian update rules over networks. In this case, agents are assumed to be boundedly rational (i.e., they fail to aggregate information in a fully Bayesian way [21]). Proposed non-Bayesian algorithms involve an aggregation step, typically consisting of weighted geometric or arithmetic average of the received beliefs [7], [22]–[25], and a Bayesian update with the locally available data [18], [26]. The aggregation step is accommodated by allowing agents to share their beliefs, or probability distributions over the hypotheses sets, with their local neighbors in the network. Note that the non-Bayesian learning setup assumes the set of hypotheses is common across the network, but the individual observations of each agent might come from different distributions. Lalitha et al. [27], Qipeng et al. [28], Shahrampour et al. [20], [29], [30], and Rahimian et al. [31] have proposed variations of the non-Bayesian approach and proved consistent, geometric and non-asymptotic convergence rates for a general class of distributed algorithms; from asymptotic analysis to non-asymptotic bounds [32], and [33], time-varying directed graphs

C.A. Uribe is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77006 USA e-mail: cauribe@rice.edu.

A. Olshevsky is with the Department of ECE and Division of Systems Engineering, Boston University, Boston, MA, 02215 USA e-mail: alexols@bu.edu.

A. Nedić is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, 85287 USA e-mail: angelia.nedich@asu.edu.

[34]. Su et al. [35] have considered adversarial agents, and transmission and node failures. Constant elasticity of substitution models [36], minimum operators [37], [38], and uncertain models [39] have also been studied. See [40] and [41] for an extended literature review.

All previous works on distributed non-Bayesian inference assume the set of hypotheses is finite. Moreover, the proof techniques provide vacuous bounds when such an assumption does not hold. In this work, we extend the centralized concentration results in [42] on non-asymptotic behaviors of Bayesian estimators on compact hypothesis sets to derive new non-asymptotic concentration results for distributed learning algorithms over networks. Our results show that, in general, the network structure will induce a transient time after which all agents learn at an independent network rate, and this rate is geometric.

The contributions of this paper (Part I) are as follows:

- We provide a variational analysis of the Bayesian posterior and derive an optimization problem for which the posterior is a step of the Stochastic Mirror Descent method.
- We use a variational interpretation to propose a distributed Stochastic Mirror Descent method for distributed learning. Moreover, we specialize the proposed algorithm to parametric models of an exponential family, which results in especially simple updates.
- We derive novel analysis methods to prove high probability non-asymptotic bounds for the convergence rate for the case of finite hypothesis sets. We show that this distributed learning algorithm concentrates the beliefs of all agents around the true parameter at an exponential rate.

The results in Part I serve as the basis for Part II of this paper series [43], where we analyze the case where the parameter spaces are compact. The objective of Part I is three-fold: 1. Motivate the general problem, 2. Analyze the specific case of inference on the exponential family of distributions, and 3. Introduce the new proof technique based on a covering of the hypothesis sets. In Part II, we extend the proof techniques to handle compact hypothesis set and provide numerical results. A subset of the problem description and a weaker set of results were presented in [44]. However, in this paper series, we extend such results with a specific treatment of the distributed inference problem for parametric estimation in the exponential family.

The rest of this paper is organized as follows. Section II introduces the problem setup, and it describes the networked observation model and the inference task. Section III presents a variational analysis of the Bayesian posterior, shows the implicit representation of the posterior as steps in a stochastic program, and extends this program to the distributed setup. Section IV specializes the proposed distributed learning protocol to the case of observation models that are members of the exponential family. Section V shows our main results about the exponential concentration of beliefs around the true parameter. Section V begins by gently introducing our techniques by proving a concentration result in the case of finite hypotheses. Finally, conclusions, open problems, and potential future work are discussed.

Notation: Random variables are denoted with upper-case letters, e.g., X, while the corresponding lower-case are used for their realizations, e.g., x. Subscripts denote time indices, and the letter k or t is generally used. Superscripts denote agent indices, and the letters i or j are used. We write $[A]_{ij}$ or a_{ij} to denote the entry of a matrix A in its i-th row and j-th column. We use A' for the transpose of a matrix A and x' for the transpose of a vector x. The complement of a set B is denoted as B^c.

II. Problem Setup

We introduce the learning problem from a centralized perspective, where all information is available at a single location. Later, we will generalize the setup to the distributed setting where only partial and distributed information is available.

Assume that we observe a sequence of independent random variables $X_1, X_2, \ldots$, all taking values in some measurable space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{X}$ is the realization space and $\mathcal{A}$ is the corresponding σ-algebra. The random variables $\{X_i\}$ are assumed identically distributed with a common *unknown* distribution P on $\mathcal{X}$, i.e., $X_k \sim P$ for all k. In addition, we have a statistical model $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ composed by a parametrized family of probability measures on the sample space $(\mathcal{X}, \mathcal{A})$, where the map $\Theta \to \mathscr{P}$ from parameter to distribution is injective. Moreover, all distributions in the model are dominated[1] by a σ-finite measure λ[2], with corresponding densities $p_\theta = dP_\theta/d\lambda$. Assume also that the model $\mathscr{P}$ is well-specified, thus *there exists a* θ^* *such that* $P_{\theta^*} = P$. The objective is to estimate θ^* based on the sequence of received observations $x_1, x_2, \ldots$. For example, given a random variable X, the maximum likelihood estimator (MLE) can be defined as

$$\hat{\theta}(X) = \arg\sup_{\theta\in\Theta} p_\theta(X) = \arg\sup_{P\in\mathscr{P}} p(X).$$

Following a Bayesian approach, the parameter is represented as a random variable ϑ on the set Θ is equipped with a σ-algebra $\mathcal{T}$ and a prior probability measure μ_0 on the measurable space $(\Theta, \mathcal{T})$, where $\mathcal{T}$ is countably generated. Moreover, we assume the existence of a probability measure Π on the product space $(\mathcal{X} \times \Theta)$ with σ-algebra $(\mathcal{A} \times \mathcal{T})$. Furthermore, the densities $p_\theta(x)$ are measurable functions of θ for any $x \in \mathcal{X}$. We then define the belief μ_k as the posterior distribution given the sequence of observations up to time k, i.e.,

$$\mu_k(B) = \Pi(\vartheta \in B \mid X_1, \ldots, X_k) = \frac{\int_B \prod_{t=1}^k p_\theta(X_t) d\mu_0(\theta)}{\int_\Theta \prod_{t=1}^k p_\theta(X_t) d\mu_0(\theta)}, \quad (1)$$

for all $B \in \mathcal{T}$ (note that we used the independence of the observations at each time step).

Assuming that all observations, up to time k, are readily available at a centralized location, under appropriate conditions, the recursive Bayesian posterior in (1) will be consistent in the sense that the beliefs μ_k will concentrate around θ^*; see [45], [46], and [47] for a formal statement. Furthermore, several authors have studied the rate at which this concentration

[1] A measure μ is dominated by (or absolutely continuous with respect to) a measure λ if $\lambda(B) = 0$ implies $\mu(B) = 0$ for every measurable set B.

[2] A positive measure defined on a σ-algebra of subsets of a set X is called σ-finite if X is the countable union of measurable sets with finite measure.

occurs, in both asymptotic and non-asymptotic regimes [42], [48], [49].

Now consider the case where there is a network of n agents observing the process $X_1, X_2, \ldots$, where X_k is now a random vector belonging to the product space $\prod_{i=1}^n \mathcal{X}^i$ and $X_k = [X_k^1, X_k^2, \ldots, X_k^n]'$. Specifically, agent i observes the sequence $X_1^i, X_2^i, \ldots$, where X_k^i is now distributed according to an unknown distribution P^i, effectively making $X_k \sim \boldsymbol{P} = \prod_{i=1}^n P^i$. The statistical model is now distributed, where each agent agent i has a private family of distributions $\mathscr{P}^i = \{P_\theta^i : \theta \in \Theta\}$ it would like to fit to the observations. However, the goal is for *all* agents to agree on a *single* θ that best explains the complete set of observations instead of their local observations only. In other words, the agents collaboratively seek to find θ^* such that $\boldsymbol{P}_{\theta^*} = \prod_{i=1}^n P_{\theta^*}^i = \prod_{i=1}^n P^i = \boldsymbol{P}$.

Agents interact over a network defined by an undirected graph $\mathcal{G} = (V, E)$, where $V = \{1, 2, \ldots, n\}$ is the set of agents and E is a set of undirected edges, i.e., $(i, j) \in E$ if and only if agents i and j can communicate with each other. We study a simple interaction model where, at each step, agents exchange their beliefs with their neighbors in the graph. Thus at every time step k, agent i will receive the sample x_k^i from X_k^i as well as the beliefs of its neighboring agents, i.e., it will receive μ_{k-1}^j for all j such that $(i, j) \in E$. We assume agents are oblivious to the network topology and the private family of distributions of other agents. Thus, fully Bayesian approaches cannot be used. *Our goal is to design a learning procedure that is both distributed and consistent. That is, we are interested in a belief update algorithm that aggregates information in a non-Bayesian manner and guarantees that the beliefs of all agents will concentrate around θ^*.*

As a motivating example, consider the problem of distributed source localization [50]. In this scenario, a network of n agents receives noisy distance measurements to a source. The sensing capabilities of each agent might be limited to a specific region. The group objective is to identify the source location jointly. There is an underlying graph that indicates which agents can exchange messages. Each agent observes signals proportional to its distance to the target. Since a target cannot be localized effectively from a single measure of the distance, agents must cooperate to have any hope of achieving proper localization. For more details on the problem, as well as simulations of the several discrete learning rules, we refer the reader to our earlier paper [32] dealing with the case when the set Θ is finite.

III. A VARIATIONAL APPROACH TO DISTRIBUTED BAYESIAN FILTERING

In this section, we observe that the posterior in (1) corresponds to an iteration of a first-order optimization algorithm, namely Stochastic Mirror Descent [51]–[54]. Closely related variational interpretations of Bayes' rule are well-known, and in particular, have been given in [55]–[57]. The specific connection to Stochastic Mirror Descent has not been noted, as far as we are aware. This connection will motivate a distributed learning method which will be the main focus of the paper.

A. Bayes' rule as Stochastic Mirror Descent

Consider the following optimization problem

$$\min_{\theta \in \Theta} F(\theta) \triangleq D_{KL}(P \| P_\theta), \tag{2}$$

where P is an unknown distribution and $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ is a parametrized family of distributions. Since P is unknown, one cannot solve (2) directly. However, we have access to realization of a random variable that is distributted according to P. Note that the optimization variable is θ. Here, $D_{KL}(P\|Q)$ is the Kullback-Leibler (KL) divergence[3] between distributions P and Q. Thus, Problem (2) consists in finding a θ in Θ such that the KL divergence between the generated distribution P_θ and P is minimized while accessing P via realizations of a random variable only.

Next, we will use the definitions of the KL divergence, to reformulate Problem (2) as a stochastic optimization problem with respect to the randomness coming from the unknown probability distribution P. The proposed estimation algorithm uses sequential observations of a random variable with distribution P. Thus, our immediate objective is to interpret the iterations of the proposed algorithm as iterates generated by a stochastic optimization method. First, note that by the definition of the KL divergence, we can rewrite Problem (2) as

$$\begin{aligned}\min_{\theta \in \Theta} D_{KL}(P\|P_\theta) &= \min_{\pi \in \Delta_\Theta} \mathbb{E}_\pi D_{KL}(P\|P_\vartheta) \qquad \text{where } \vartheta \sim \pi \\ &= \min_{\pi \in \Delta_\Theta} \mathbb{E}_\pi \mathbb{E}_P \left[-\log \frac{dP_\vartheta(X)}{dP(X)} \right] \qquad \text{where } \vartheta \sim \pi, X \sim P,\end{aligned}$$

and Δ_Θ is the set of all possible distributions on the parameter space Θ, and $dP_\vartheta(X)/dP(X)$ is the Radon–Nikodym derivative of P_ϑ with respect to P. Since the distribution P does not depend on ϑ, it follows that

$$\begin{aligned}\arg\min_{\pi \in \Delta_\Theta} \mathbb{E}_\pi \mathbb{E}_P \left[-\log \frac{dP_\vartheta(X)}{dP(X)} \right] &= \arg\min_{\pi \in \Delta_\Theta} \mathbb{E}_\pi \mathbb{E}_P \left[-\log p_\vartheta(X) \right] \\ &= \arg\min_{\pi \in \Delta_\Theta} \mathbb{E}_P \mathbb{E}_\pi \left[-\log p_\vartheta(X) \right]. \end{aligned} \tag{3}$$

The equality in (3), where we exchange the order of the expectations, follows from the Fubini-Tonelli theorem [58, Page 233]. If θ^* minimizes (2), then a distribution π^* which puts all the mass on θ^* (i.e. $\pi^*(\vartheta = \theta^*) = 1$) minimizes (3).

The difficulty in evaluating the objective function (3) lies in the fact that the distribution P is unknown. A generic approach to solving such problems is using algorithms from stochastic approximation methods. The objective is minimized by constructing a sequence of gradient-based iterates whereby the true gradient of the objective (which is not available) is replaced with a gradient sample available at a given time.

A particular method that is relevant for the solution of stochastic programs as in (3) is the *stochastic mirror descent* method [51], [52], [59], [60]. In particular, recall that the mirror descent method to find the minimum of a function $f(x)$ performs the update

$$x_{k+1} \in \arg\min \left\{ \nabla f(x_k)' x + \frac{1}{\alpha_k} D(x, x_k) \right\},$$

[3] $D_{KL}(P\|Q)$ between distributions P and Q (with P dominated by Q) is defined to be $D_{KL}(P\|Q) = -\mathbb{E}_P \left[\log dQ/dP \right]$.

where $D(\cdot,\cdot)$ is a Bregman divergence. Later on, we will study the specific case where the used Bregman divergence is the KL divergence. Moreover, note that (3) is linear in π. Thus, we have that $d/d\pi \left\langle \mathbb{E}_P\left[-\log p_\vartheta(X)\right], \pi\right\rangle = \mathbb{E}_P\left[-\log p_\vartheta(X)\right]$. Finally, we use the stochastic approximation provided by the current sample x_{k+1} of X. Therefore, the stochastic mirror descent approach constructs a sequence of densities $\{d\mu_k\}$, as

$$d\mu_{k+1} = \underset{\pi\in\Delta_\Theta}{\arg\min}\left\{\langle -\log p_\theta(x_{k+1}), \pi\rangle + \frac{1}{\alpha_k} D_w(\pi, d\mu_k)\right\}, \tag{4}$$

where $\alpha_k > 0$ is the step-size, the inner product is defined as $\langle p, q\rangle = \int_\Theta p(\theta)q(\theta)d\sigma$. Moreover, $D_w(x, x_k)$ is a (functional) Bregman divergence [61] associated with a strictly convex, and twice continuously Fréchet differentiable functional $w : L^p(\nu) \to \mathbb{R}$, i.e., $D_w(x,z) = w(x) - w(z) - \delta w[z; x-z]$, where $\delta w[z; x-z]$ is the Fréchet derivative of w at z in the direction of $x - z$. If we choose $w(x) = \int x(\nu)\log x(\nu) d\nu$ as the distance-generating function, then the corresponding Bregman distance is the Kullback-Leibler (KL) divergence D_{KL}. Additionally, by selecting $\alpha_k = 1$, the solution to Problem (4) can be computed explicitly, where for each $\theta \in \Theta$,

$$d\mu_{k+1}(\theta) \propto p_\theta(x_{k+1}) d\mu_k(\theta),$$

which is the posterior distribution as defined in (1) (a formal proof of this assertion is a special case of Proposition 1 shown later in the paper).

We have just shown how Bayes rule, i.e., the posterior computation, can be viewed as an instance of mirror descent with a stochastic approximation for a particular choice of Bregman function; in the following subsection, we show how this interpretation leads to a natural algorithm in the distributed Bayesian posterior.

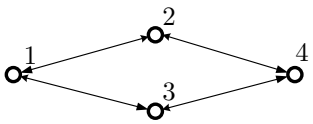
B. Distributed Stochastic Mirror Descent

Now, consider the distributed problem where the network of agents want to collectively solve the following optimization problem

$$\min_{\theta\in\Theta} F(\theta) \triangleq D_{KL}\left(\boldsymbol{P}\|\boldsymbol{P}_\theta\right) = \sum_{i=1}^n D_{KL}(P^i\|P^i_\theta). \tag{5}$$

Recall that the distribution $\boldsymbol{P}$ is unknown (though, of course, agents gain information about it by observing samples from $X^i_1, X^i_2, \ldots$ and interacting with other agents) and that $\mathscr{P}^i$ containing all the distributions P^i_θ is a private family of distributions and is only available to agent i. A distributed method should guarantee all agents in the network concentrate their beliefs around θ^* which is defined as a solution of (5).

For example, consider a group of 4 agents connected over a network as shown in Figure 1.



$$A = \begin{bmatrix} 2/3 & 1/6 & 0 & 1/6 \\ 1/6 & 2/3 & 1/6 & 0 \\ 0 & 1/6 & 2/3 & 1/6 \\ 1/6 & 0 & 1/6 & 2/3 \end{bmatrix}$$

Fig. 1: A network of 4 agents.

Furthermore, assume that each agent is observing a Bernoulli random variable such that $X^1_k \sim \text{Bern}(0.2)$, $X^2_k \sim \text{Bern}(0.4)$, $X^3_k \sim \text{Bern}(0.6)$ and $X^4_k \sim \text{Bern}(0.8)$. In this case, the parameter space is $\Theta = [0,1]$. Thus, the objective is to collectively find a parameter θ^* that best explains the joint observations in the sense of (5), i.e.,

$$\begin{aligned}\min_{\theta\in[0,1]} F(\theta) &= \sum_{j=1}^4 D_{KL}(\text{Bern}(\theta^j)\|\text{Bern}(\theta)) \\ &= \sum_{j=1}^4 \left(\theta\log\frac{\theta}{\theta^j} + (1-\theta)\log\frac{1-\theta}{1-\theta^j}\right)\end{aligned}$$

where $\theta^1 = 0.2$, $\theta^2 = 0.4$, $\theta^3 = 0.6$ and $\theta^4 = 0.8$. The optimal solution is $\theta^* = 0.5$ by the first-order optimality conditions or by exploiting symmetries in the objective function.

We propose the following algorithm as a distributed version of the stochastic mirror descent for the solution of (5):

$$\begin{aligned} d\mu^i_{k+1} = \underset{\pi\in\Delta_\Theta}{\arg\min}\left\{\langle -\log p^i_\theta(x^i_{k+1}), \pi\rangle + \sum_{j=1}^n a_{ij} D_{KL}(\pi\|d\mu^j_k)\right\} \\ \text{where } \theta\sim\pi, \end{aligned} \tag{6}$$

with $a_{ij} > 0$ denoting the weight that agent i assigns to beliefs coming from its neighbor j. Specifically, $a_{ij} > 0$ if $(i,j) \in E$ or $j = i$, and $a_{ij} = 0$ if $(i,j) \notin E$. For example, Figure 1 shows a set of metropolis weights for this network is given by the matrix $A = (a_{ij})$. Problem (6) has a closed form solution. In particular, the posterior density at each $\theta \in \Theta$ is given by

$$d\mu^i_{k+1}(\theta) \propto p^i_\theta(x^i_{k+1}) \prod_{j=1}^n (d\mu^j_k(\theta))^{a_{ij}},$$

or equivalently, the belief on a measurable set B of an agent i at time $k+1$ is

$$\mu^i_{k+1}(B) \propto \int_B p^i_\theta(x^i_{k+1}) \prod_{j=1}^n (d\mu^j_k(\theta))^{a_{ij}}. \tag{7}$$

We state the correctness of this claim in the following proposition.

Proposition 1: Assume that for an agent i the set of weights $(a_{ij})_{j=1}^n$ form a stochastic vector, and the matrix defined by all weights (a_{ij}) is symmetric. Then, the probability measure μ^i_{k+1} over the set Θ defined in (7) is a solution of Problem (6) for every agent i. Moreover, the generated sequence coincides, almost everywhere, with the update of the distributed stochastic mirror descent algorithm applied to Problem (5).

The proof of Proposition 1 can be found in the Appendix. We remark that the update (7) can be viewed as a two-step process: first, every agent constructs an aggregate belief using a weighted geometric average of its own belief and the beliefs of its neighbors, and then each agent performs a Bayes' update using the aggregated belief as a prior. We note that similar arguments in the context of distributed optimization have been proposed in [54], [62] for general Bregman distances. In the case when the number of hypotheses is finite, variations on this update rule were previously analyzed in [27], [29], [32].

To summarize, we have interpreted Bayes' rule as an instance of Stochastic Mirror Descent. We have shown how this interpretation motivates a distributed update rule. The following section discusses explicit forms of this update rule for parametric models coming from exponential families.

IV. Cooperative Inference for Exponential Families

We begin with the observation that, for a general class of models $\{\mathscr{P}^i\}$, the direct computation of the posterior beliefs μ^i_{k+1} is intractable. Indeed, computing μ^i_{k+1} requires the exact computation of an integral of the form

$$\int_\Theta p^i_\theta(x^i_{k+1}) \prod_{j=1}^n (d\mu^j_k(\theta))^{a_{ij}}. \tag{8}$$

There is an entire area of research called *variational Bayes' approximations* dedicated to efficiently approximating integrals that appear in such context [63]–[65].

This section shows that for exponential family [66], [67] there are closed-form expressions for the posterior beliefs generated by the proposed distributed inference algorithm. The exponential family of distributions is sufficiently general as many of the commonly used distributions are members of the exponential family, e.g., Normal, exponential, gamma, chi-square, beta, Dirichlet, Bernoulli, categorical, Poisson, Wishart, inverse Wishart, and geometric, among others. Also, exponential families have additional properties that make them useful for statistical analysis such as sufficient statistics exist, explicit conjugate priors, closed-form posterior predictive distributions, and optimal approximations in the mean-field approximation in variational Bayes [68]–[70].

Definition 1: The exponential family, for a parameter $\theta = [\theta^1, \theta^2, \dots, \theta^s]'$, is the set of probability distributions whose density can be represented as

$$p_\theta(x) = H(x) \exp(M(\theta)' T(x)),$$

for specific functions $H(\cdot)$, $M(\cdot)$ and $T(\cdot)$ where $M(\theta) = [M(\theta^1), M(\theta^2), \dots, M(\theta^s)]'$ depends on the density distribution of the parameters and $T(\cdot)$ depends on the observations.

For example, consider a Gaussian distribution parametrized by its mean θ with known variance σ^2. Then, it holds that

$$\begin{aligned} p_\theta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) \\ &= \underbrace{\frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}}_{H(x)} \exp\left(\underbrace{\begin{bmatrix}\theta & \theta^2\end{bmatrix}}_{M(\theta)} \underbrace{\begin{bmatrix}\frac{x}{\sigma^2} \\ -\frac{1}{2\sigma^2}\end{bmatrix}}_{T(x)} \right). \end{aligned} \tag{9}$$

Among the exponential family members, one can find distributions such as Gaussian, Poisson, Exponential, Gamma, Bernoulli, and Beta, among others [71]. In our case, we will take advantage of the existence of *conjugate priors* for all members of the exponential family. The definition of the conjugate prior is given as follows.

Definition 2: Assume that the prior distribution p on a parameter space Θ belongs to the exponential family. Then, the distribution p is referred to as the *conjugate prior* for a likelihood function $p_\theta(x)$ if the posterior distribution $p(\theta|x) \propto p_\theta(x)p(\theta)$ is in the same family as the prior.

Definition 2 implies that if the belief density at some time k is a conjugate prior for our likelihood model, then our belief at time $k+1$ will be of the same class as our prior. For example, if a likelihood function follows a Gaussian form, having a Gaussian prior will produce a Gaussian posterior. This property simplifies the structure of the belief update procedure since we can express the evolution of the beliefs generated by (7) by the evolution of the natural parameters of the member of the exponential family it belongs to. Naturally, by induction, if the prior belief at time $k=0$ is a conjugate prior to the likelihood function, the beliefs for all $k>0$ will belong to the same exponential family.

In the same way that a Gaussian likelihood function can be represented in its canonical form as in (9), we can find such representation for the belief density. Note, however, that in this case, the sample space is not $\mathcal{X}$ as in the likelihood function, but Θ because the belief is a distribution over Θ. Moreover, we will require some parametric characterization. Particularly we can write a belief density as

$$p_\chi(\theta) = f(\chi) \exp(M(\theta)'\chi),$$

where M is a function of the parameter space for $\theta \in \Theta$ and χ is the set of *natural parameters*, which is a parametric characterization of the belief density.

Going back to the example in (9), assume that our prior is a Gaussian distribution on θ with mean $\hat{\theta}$ and variance $\hat{\sigma}^2$, then $\chi' = [\hat{\theta} \;\; \hat{\sigma}^2]'$ and

$$\begin{aligned} p_\chi(\theta) &= \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{(\theta-\hat{\theta})^2}{2\hat{\sigma}^2}\right) \\ &= \underbrace{\frac{\exp\left(-\frac{\hat{\theta}^2}{2\hat{\sigma}^2}\right)}{\sqrt{2\pi\hat{\sigma}^2}}}_{f(\chi)} \exp\left(\underbrace{\begin{bmatrix}\theta & \theta^2\end{bmatrix}}_{M(\theta)} \underbrace{\begin{bmatrix}\frac{\hat{\theta}}{\hat{\sigma}^2} \\ -\frac{1}{2\hat{\sigma}^2}\end{bmatrix}}_{\chi} \right). \end{aligned} \tag{10}$$

It follows from (10) that $f(\chi) = \frac{\exp\left(-\frac{\hat{\theta}^2}{2\hat{\sigma}^2}\right)}{\sqrt{2\pi\hat{\sigma}^2}}$, $M(\theta) = \begin{bmatrix}\theta, \theta^2\end{bmatrix}$, and $\chi' = \begin{bmatrix}\hat{\theta}/\hat{\sigma}^2, -1/2\hat{\sigma}^2\end{bmatrix}$.

Then, it can be shown that the posterior distribution, given some observation x, has the same exponential form as the prior, with updated parameter $\bar{\chi} = \chi + T(x)$ as follows:

$$p_{\bar{\chi}}(\theta|x) \propto p_\theta(x) p_\chi(\theta|x). \tag{11}$$

Particularly, for the example in (9) and (10), the posterior distribution is still Gaussian.

We will now exploit the structure of the exponential family of distributions to reformulate (7) into an easy-to-implement algorithm in terms of the parametric representation of the beliefs for each agent.

Initially, consider that the set of agents have a belief at time k in the form of a distribution over the parameter space that

is a member of the exponential family. That is, assume that each agent i has a belief over the parameters θ such that

$$d\mu_k^i(\theta) \propto \exp\left(M(\theta)'\chi_k^i\right),$$

then, according to the first step in (7), an agent i needs to compute the weighted geometric average of the beliefs of its neighbors including its own. Given the parametrization in the exponential family, it holds that,

$$\prod_{j=1}^{n}\Big(d\mu_k^j(\theta)\Big)^{a_{ij}} \propto \exp\big(M(\theta)'\sum_{j=1}^{n}a_{ij}\chi_k^j\big).$$

If all agents have beliefs in the same exponential family and they are conjugate priors to their corresponding likelihood functions, then we can write the posterior of agent i as

$$\begin{aligned} d\mu_{k+1}^i(\theta) &\propto \exp\big(M(\theta)'\sum_{j=1}^{n}a_{ij}\chi_k^j\big)p_M^i(x_{k+1}^i) \\ &= \exp\big(M(\theta)'\big(\sum_{j=1}^{n}a_{ij}\chi_k^j + T^i(x_{k+1}^i)\big)\big). \end{aligned}$$

Note that the above relation defines a new distribution in the same exponential family with a new parameter $\chi_{k+1}^i = \sum_{j=1}^n a_{ij}\chi_k^j + T^i(x_{k+1}^i)$. As an immediate conclusion, it follows that for distributed inference problems when the observation models are members of the exponential family, one can always construct a set of beliefs using prior conjugates, and (7) simplifies to updates in the parameters of the exponential family, as shown by the following proposition.

Proposition 2: Assume the belief density $d\mu_k^i$ at time k has an exponential form with natural parameters χ_k^i for all $1 \leq i \leq n$, and that these densities are conjugate priors of the likelihood models p_θ^i. Then, the belief density of agent i at time $k+1$, as computed in (7), has the same form as the beliefs at time k with the natural parameters given by

$$\chi_{k+1}^i = \sum_{j=1}^{n}a_{ij}\chi_k^j + T^i(x_{k+1}^i). \tag{12}$$

A. Examples

In this subsection, we explicitly state the general distributed algorithm in (12) presented in Proposition 2 for several distributed parameter estimation problems. Mainly, we explicitly write the definition of the vector $T^i(x_k^i)$ and χ_k^i, from which the parameters of the current beliefs for each agent can be computed.

1) Distributed Gaussian Filter with unknown mean and known variance: Assume each agent in the network observes a signal of the form $X_k^i = \theta^i + \epsilon_k^i$, where θ^i is finite and unknown scalar quantity, while $\epsilon^i \sim \mathcal{N}(0, 1/\tau^i)$ is a zero mean Gaussian noise with precision $\tau^i = 1/(\sigma^i)^2$ known only by agent i. The objective of the network is to agree on a single θ^* that solves (5).

In this case, the likelihood models, the prior and the posterior are Gaussian distributions. Thus, if the beliefs of the agents at time k are Gaussian, i.e., $\mu_k^i = \mathcal{N}(\theta_k^i, 1/\tau_k^i)$ for all $i = 1\ldots, n$, then their beliefs at time $k+1$ are also Gaussian. In particular, they are given by $\mu_k^i = \mathcal{N}(\theta_k^i, 1/\tau_k^i)$ for all $i = 1\ldots, n$, with

$$M(\theta) = \begin{bmatrix}\theta \\ \theta^2\end{bmatrix}, \qquad T^i(x_k^i) = \begin{bmatrix}x_k^i\tau^i \\ -\frac{1}{2}\tau^i\end{bmatrix}, \qquad \chi_k^i = \begin{bmatrix}\theta_k^i\tau_k^i \\ -\frac{1}{2}\tau_k^i\end{bmatrix}.$$

We note that this specific setup is known as Gaussian Learning and has been studied in [72], [73], where the expected parameter estimator is shown to converge at an $O(1/k)$ rate.

2) Distributed Gaussian Filter with unknown variance and known mean: In this case, the agents want to cooperatively estimate the value of a variance which is the parameter for (5). Specifically, each agent i observes a realization of the random variable $X_k^i = \theta^i + \epsilon_k^i$, with $\epsilon_k^i \sim \mathcal{N}(0, 1/\tau^i)$, where θ^i is known and τ^i is unknown. The beliefs of all agents are chosen to be a Gamma distribution $\mu_k^i = \text{Gamma}(\alpha_k^i, \beta_k^i)$ and it follows that

$$M(\tau) = \begin{bmatrix}\tau \\ \log\tau\end{bmatrix}, T^i(x_k^i) = \begin{bmatrix}-\frac{1}{2}(x_k^i-\theta^i)^2 \\ -\frac{1}{2}\end{bmatrix}, \chi_k^i = \begin{bmatrix}-\beta_k^i \\ -(\alpha_k^i-1)\end{bmatrix}.$$

3) Distributed Gaussian Filter with unknown mean and variance: We have considered the cases when either the mean or the variance is known in the preceding examples. Here, we will assume that both the mean and the variance are unknown and need to be estimated. Explicitly, we still have noise observations $X_k^i = \theta^i + \epsilon_k^i$, with $\epsilon_k^i \sim \mathcal{N}(0, 1/\tau^i)$. We are going to assume all agents have beliefs that follow the Gaussian-Gamma distribution, i.e. $\mu_k^i = \text{GaussianGamma}(\theta_k^i, \lambda_k^i, \alpha_k^i, \beta_k^i)$ for $i = 1, \cdots, n$. Moreover, the it holds that

$$M(\theta,\tau)= \begin{bmatrix}\log\tau \\ \tau \\ \tau\theta \\ \tau\theta^2\end{bmatrix}, T^i(x_k^i)= \begin{bmatrix}\frac{1}{2} \\ -\frac{1}{2}(x_k^i)^2 \\ x_k^i \\ -\frac{1}{2}\end{bmatrix}, \chi_k^i= \begin{bmatrix}\alpha_k^i-\frac{1}{2} \\ -\frac{1}{2}\lambda_k^i(\theta_k^i)^2-\beta_k^i \\ \lambda_k^i\theta_k^i \\ -\frac{1}{2}\lambda_k^i\end{bmatrix}.$$

4) Distributed Bernoulli Filter: Here, each of the agents receives private observations of the form $X_k^i \sim \text{Bernoulli}(p^i)$, with p^i unknown. In order to estimate the network-wide parameter, each agent constructs a sequence of beliefs following a Beta distribution, i.e. $\mu_k^i = \text{Beta}(\alpha_k^i, \beta_k^i)$. Then, (12) updates its parameters. Moreover, it holds that

$$M(p) = \begin{bmatrix}\log p \\ \log(1-p)\end{bmatrix}, \; T^i(x_k^i) = \begin{bmatrix}x_k^i \\ 1-x_k^i\end{bmatrix}, \; \chi_k^i = \begin{bmatrix}\alpha_k^i \\ \beta_k^i\end{bmatrix}.$$

5) Distributed Poisson Filter: Similarly as before, we consider an observation model where each agent i receives realization of a Poisson random variable with unknown parameter λ^i, i.e., $X_k^i \sim \text{Poisson}(\lambda^i)$ for all i. The conjugate prior of a Poisson likelihood model is the Gamma distribution. Thus, if at time k the beliefs of each agent i are given by $\mu_k^i = \text{Gamma}(\alpha_k^i, \beta_k^i)$. Moreover, it holds that

$$M(\lambda) = \begin{bmatrix}\log\lambda \\ \lambda\end{bmatrix}, \; T^i(x_k^i) = \begin{bmatrix}x_k^i \\ -1\end{bmatrix}, \; \chi_k^i = \begin{bmatrix}\alpha_k^i - 1 \\ -\beta_k^i\end{bmatrix}.$$

6) Distributed Exponential Filter: As a final example, we consider an observation model where each agent i receives realization of an Exponential random variable with unknown rate λ^i, i.e., $X_k^i \sim \text{Exponential}(\lambda^i)$ for all i. The conjugate prior of an Exponential likelihood model is the Gamma distribution. Thus, if at time k the beliefs of each agent i are given by $\mu_k^i = \text{Gamma}(\alpha_k^i, \beta_k^i)$. Moreover, it holds that

$$M(\lambda) = \begin{bmatrix} \lambda \\ \log \lambda \end{bmatrix}, \; T^i(x_k^i) = \begin{bmatrix} -1 \\ x_k^i \end{bmatrix}, \; \chi_k^i = \begin{bmatrix} \alpha_k^i - 1 \\ -\beta_k^i \end{bmatrix}.$$

V. Belief Concentration Rates

We now turn to the presentation of our main results about the rate at which beliefs generated by (7) concentrate around the true parameter θ^*. We will break up our analysis into two cases. Initially, Part I of this paper series will focus on when Θ is a finite set and will prove a concentration rate on the beliefs on a Hellinger ball around the optimal hypothesis. The case when Θ is a finite set has been previously studied in [27], [29], [32], where similar geometric concentration results for distributed learning have been shown. However, we take a fundamentally different proof approach that will allow us to gently introduce the techniques we will use later when we turn our attention to the case when Θ is a compact subset of $\mathbb{R}^d$. We analyze the case of compact hypothesis sets in Part II of this paper series [43]. Our proof techniques use concentration arguments for beliefs on Hellinger balls from the recent work in [42] which, in turn, builds on [74].

We begin with two subsections focusing on background information, definitions, and assumptions.

A. Background: Hellinger Distance and Coverings

The *squared* Hellinger distance between two probability distributions P and Q is given by,

$$h^2(P,Q) = \frac{1}{2}\int \left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda, \qquad (13)$$

where P and Q are dominated by λ. Moreover, the Hellinger distance satisfies the property that $0 \leq h(P,Q) \leq 1$.

We equip the set of all probability distributions $\mathscr{P}$ over the parameter set with the Hellinger distance to obtain the *metric* space $(\mathscr{P}, h)$. The metric space induces a topology, where we can define an open ball $\mathcal{B}_r(\theta)$ with a radius $r \in (0,1)$ centered at a point $\theta \in \Theta$, which we use to construct a special covering of subsets $B \subset \mathscr{P}$. Recall that (13) defines the squared Hellinger distance h^2, rather than h.

Definition 3: Define an n-Hellinger ball of radius r centered at θ as

$$\mathcal{B}_r(\theta) = \left\{ \hat{\theta} \in \Theta \;\middle|\; \frac{1}{n}\sum_{i=1}^n h^2\left(P_\theta^i, P_{\hat{\theta}}^i\right) \leq r^2 \right\}.$$

Additionally, when no center is specified, it should be assumed that it refers to θ^*, i.e. $\mathcal{B}_r = \mathcal{B}_r(\theta^*)$.

Given an n-Hellinger ball of radius r, we will use the following notation for a covering of its complement $\mathcal{B}_r^c$. Specifically, we are going to express $\mathcal{B}_r^c$ as the union of finite disjoint and concentric annuli. Let $r \in (0,1)$ and $\{r_l, l = 1, \ldots, L\}$ be a finite strictly decreasing sequence such that $r_1 = 1$ and $r_L = r$, and express the set $\mathcal{B}_r^c$ as the union of annuli generated by the sequence $\{r_l\}$ as

$$\mathcal{B}_r^c \subseteq \bigcup_{l=1}^{L-1} \mathcal{F}_l,$$

where $\mathcal{F}_l = \mathcal{B}_{r_l} \setminus \mathcal{B}_{r_{l+1}}$. Note that the above relation holds true because the maximum Hellinger distance is 1.

B. Background: Assumptions on the Network and Mixing Weights

Naturally, we need some assumptions on the matrix A. For one thing, the matrix A has to be "compatible" with the underlying graph, in that information from node i should not affect node j if there is no edge from i to j in $\mathcal{G}$. At the other extreme, we want to rule out the possibility that A is the identity matrix, which in terms of (7) means nodes do not talk to their neighbors. Formally, we make the following assumption.

Assumption 1: The undirected graph $\mathcal{G} = (V, E)$, where $V = 1, \ldots, n$ is the set of vertices and E is the set of undirected edges, and matrix A are such that:

(a) A is symmetric and doubly-stochastic with $[A]_{ij} = a_{ij} > 0$ for $i \neq j$ if and only if $(i,j) \in E$.
(b) A has positive diagonal entries, $a_{ii} > 0$ for all $i \in V$.
(c) The graph $\mathcal{G}$ is connected.

Assumption 1 is common in the distributed optimization literature. The construction of a set of weights satisfying Assumption 1 can be done in a distributed way, for example, by choosing the so-called "lazy Metropolis" matrix, which is a stochastic matrix given by

$$a_{ij} = \begin{cases} \frac{1}{2\max\{d^i+1, d^j+1\}} & \text{if } (i,j) \in E, \\ 0 & \text{if } (i,j) \notin E, \end{cases}$$

where d^i is the degree (the number of neighbors) of node i. Note that although the above formula only gives the off-diagonal entries of A, it uniquely defines the entire matrix (the diagonal elements are uniquely defined via the stochasticity of A). To choose the weights corresponding to a lazy Metropolis matrix, agents will need to spend an additional round at the beginning of the algorithm broadcasting their degrees to their neighbors.

Assumption 1 can be seen to guarantee that $A^k \to (1/n)\mathbf{1}\mathbf{1}^T$ where $\mathbf{1}$ is the vector of all ones. We will use the following result based on [29] and [32], that provides convergence rate for the difference $|A^k - (1/n)\mathbf{1}\mathbf{1}^T|$.

Lemma 3: Let Assumption 1 hold, then the matrix A satisfies the following relation:

$$\sum_{t=1}^k \sum_{j=1}^n \left| \left[A^{k-t}\right]_{ij} - \frac{1}{n} \right| \leq \frac{4\log n}{1-\delta} \qquad \text{for } i = 1, \ldots, n,$$

where $\delta = \max\{|\lambda_n(A)|, |\lambda_2(A)|\}$ is the maximum eigenvalue of the matrix A, and $1-\delta$ is refereed as the *spectral gap*. For an arbitrary doubly stochastic matrix A, it holds that $\delta = 1 - \eta/4n^2$ with η being the smallest positive entry of the matrix A. Furthermore, if A is a lazy Metropolis matrix associated with the graph $\mathcal{G}$, then $\delta = 1 - 1/\mathcal{O}(n^2)$.

C. Concentration Analysis for Finite Hypothesis Sets

We now turn to prove a concentration result when the set Θ of hypotheses is finite. We will show exponential convergence of beliefs on a Hellinger Ball around the true hypothesis θ^*. The primary purpose of this analysis is to gently introduce the techniques that will be used later for the case of a compact set of hypotheses. If the number of hypotheses is finite, (7)

can be written in a simpler form for discrete beliefs over the parameter space Θ as

$$\mu^i_{k+1}(\theta) \propto p^i_\theta(x^i_{k+1}) \prod_{j=1}^{n} (\mu^j_k(\theta))^{a_{ij}}. \tag{14}$$

We will fix the radius r, and our goal will be to prove a concentration result for a Hellinger ball of radius r around the optimal hypothesis θ^*. We start by partitioning the complement of this ball, i.e., $\mathcal{B}^c_r$, as described above, into the annuli $\mathcal{F}_l$. We introduce the notation $\mathcal{N}_{r_l}$ to denote the number of hypotheses within the annulus $\mathcal{F}_l$.

The Hellinger distance allows us to use the Hellinger affinity between two distributions Q and P, defined as $\rho(Q,P) = 1 - h^2(Q,P)$. We are now ready to state our first result as a lemma that bounds the concentration of aggregated log-likelihood ratios.

Lemma 4: Let Assumption 1 hold. Given a set of independent random variables $\{X^i_t\}$ such that $X^i_t \sim P^i$ for $i = 1,\dots,n$, and $t = 1,\dots,k$, a set of distributions $\{Q^i\}$ where P^i dominates Q^i, then for all $y \in \mathbb{R}$,

$$\mathbb{P}\left[\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij} \log \frac{dQ^j}{dP^j}(X^j_t) \geq y\right] \leq \exp\left(-\frac{y}{2} + \frac{4\log n}{1-\delta} - \frac{k}{n}\sum_{j=1}^{n} h^2(Q^j,P^j)\right).$$

The proof of Lemma 4 can be found in the Appendix. We are now ready to state our first main result, which bounds the concentration of (14) around the optimal hypothesis for a finite hypothesis set Θ. The following theorem shows that all agents' beliefs will concentrate around the Hellinger ball $\mathcal{B}_r$ at an exponential rate.

Theorem 5: Let Assumption 1 hold, $\sigma \in (0,1)$ be a desired probability tolerance, $\omega \in (0,1)$, $r \in (0,1)$, and $\{r_l, l = 1,\dots,L\}$ be a finite strictly decreasing sequence such that $r_1 = 1$ and $r_L = r$. Then, the belief sequences $\{\mu^i_k\}$, for $i = 1,\dots,n$, generated by (14), with identical initial beliefs on all agents such that $\mu^j_0(\theta) = \mu^i_0(\theta)$ for all $i,j = 1,\dots,m$ and $\theta \in \Theta$, have the following property: with probability $1-\sigma$,

$$\mu^i_{k+1}(\mathcal{B}_r) \geq 1-\omega \quad \forall i \text{ and } k \geq N,$$

where

$$N = \inf\left\{t \geq 1 \middle| \exp\left(\frac{4\log n}{1-\delta}\right)\sum_{l=1}^{L-1}\mathcal{N}_{r_l}\exp\left(-t r^2_{l+1}\right) < \sigma\sqrt{\omega\mu_0(\theta^*)}\right\},$$

$\mathcal{N}_{r_l}$ is the number of hypotheses within the annulus $\mathcal{F}_l = \mathcal{B}_{r_l} \setminus \mathcal{B}_{r_{l+1}}$, and $\delta = 1 - \eta/n^2$, where η is the smallest positive element of the matrix A.

The proof of Theorem 5 can be found in the Appendix. In Theorem 5, note that N indicates the time required for the beliefs on the ball $\mathcal{B}_r$ around the true hypothesis θ^* to be at least $1-\omega$. Moreover, N is a function of the probability tolerance σ, the desired concentration parameter ω, and the number of agents n. The time N can also be interpreted as a transient time required for mixing of beliefs among the agents before the impact of the network disappears. N will be larger with smaller values of ω and σ, for larger values of n, or δ close to 1. The smaller the radius r, the larger the number of hypotheses outside the Hellinger ball of radius r, making the parameters $\mathcal{N}_{r_l}$ larger, which implies larger N.

Note that in general, the belief concentration rate described in Theorem 5 depends on the geometry of the hypotheses set and how they are distributed on the parameter space. The next Corollary describes the scenario where the sequence $\{r_l\}$ is such that $L = 2$, so $r_1 = 1$ and $r_2 = r$.

Corollary 6: Let Assumption 1 hold, and let $\sigma \in (0,1)$ be a desired probability tolerance. Then, the belief sequences $\{\mu^i_k\}$, $i = 1,\dots,n$ that are generated by (14), with identical initial beliefs on all agents such that $\mu^i_0(\theta^*) > \epsilon$ for all i, have the following property: for any radius $r \in (0,1)$ with probability $1-\sigma$, $\mu^i_k(\mathcal{B}_r) \geq 1-\omega$, for $k \geq \frac{1}{r^2}\left(\log\left(\mathcal{N}/(\sigma\sqrt{\omega\mu_0(\theta^*)})\right) + \frac{4\log n}{1-\delta}\right)$ where $\mathcal{N}$ is the number of hypotheses outside $\mathcal{B}_r$ and δ is defined in Lemma 3.

D. Discussion and Comparison with Previous Approaches

Non-asymptotic belief concentration rates for non-Bayesian learning has been previously studied in [27], [29], [32]. In this subsection, we provide some discussion and comparison with the result from Theorem 5, and Corollary 6 respectively. We start by recalling a general form of the main result from [27], [29], [32].

Theorem 7 (Theorem 2 from [32]): Let Assumptions 1 hold and let $\sigma \in (0,1)$. The update rule (14), with positive and uniform initial belief on all hypotheses, has the following property: there is an integer $\boldsymbol{N}(\sigma)$ such that, with probability $1-\rho$, for all $k \geq \boldsymbol{N}(\sigma)$ and for all $\theta_v \notin \Theta^*$, we have

$$\mu^i_k(\theta_v) \leq \exp\left(-\frac{k}{2}\gamma_2 + \gamma^i_1\right) \quad \text{for all } i = 1,\dots,n,$$

where $\boldsymbol{N}(\sigma) \triangleq \left\lceil \frac{1}{\gamma^2_2} 8\left(\log\alpha\right)^2 \log\frac{1}{\sigma}\right\rceil$, and

$$\gamma^i_1 \triangleq \frac{12\log n}{1-\delta}\log\frac{1}{\alpha}, \quad \gamma_2 \triangleq \frac{1}{n}\min_{\theta_v \notin \Theta^*}\sum_{i=1}^{n} D_{KL}(P^i \| P^i_{\theta_v}),$$

where α is a positive lower bound on the likelihood functions.

For simplicity of presentation, we will focus our comparison with Corollary 6. Initially, note that Corollary 6 indicates the concentration of beliefs on an n-Hellinger ball of radius r around the optimal hypotheses, whereas Theorem 7 shows that the beliefs on the non-optimal hypotheses will decay to zero. These two statements are equivalent if the optimal hypothesis is unique, and the n-Hellinger ball of radius r contains only one hypothesis. Moreover, the rate at which such concentrations occur is exponential in the number of iterations for both cases. However, the rate in Corollary 6 is given by the radius r, whereas in Theorem 7 is given by the distance between the optimal and second-best hypotheses. These two statements seem equivalent. In Corollary 6 the distance is measured in terms of Hellinger distances, which are naturally upper bounded by 1. In Theorem 7, the Kullback-Leibler divergence which is an upper bound for the squared Hellinger distance.

Such relation follows naturally from Pinsker's inequality for Hellinger distances [75, Lemma 2.4] where $1/2D_{KL}(P\|Q) \geq h^2(P,Q)$. This weakness of the proposed method might be explained as Problem (5) involved KL divergences. Nevertheless, this is a trade-off for a more general analysis that will work on compact hypotheses spaces. We believe this is an artifact of the proof. Removing such construction is out of the scope of this paper and left for future work. The belief concentrations for both results happen after a time proportional to a term that depends on the network topology. They are equal up to a constant factor of 3. Finally, an essential difference to previous works is that Corollary 6 removes the lower bounded likelihood assumption in Theorem 7, i.e., the term α. The KL divergence grows unboundedly when the reference distribution is not absolutely continuous with respect to the other one. Corollary 6 effectively allows extending previous results to discrete distributions with different support. In both cases, the dependency on the high probability bound is only logarithmic.

VI. Conclusions

We have proposed an algorithm for cooperative distributed non-Bayesian learning over networks. Our algorithm may be viewed as a distributed version of Stochastic Mirror Descent applied to the problem of minimizing the sum of Kullback-Leibler divergences. Our results show non-asymptotic geometric convergence rates for the beliefs concentration around the true hypothesis. Particularly in Part I, we provide an extensive application case of study for observational models in the exponential family of probability distributions. Moreover, we have developed a new belief concentration analysis for the case of finite hypotheses. Part II of this paper series [43] extends this analysis to the compact hypotheses set case.

Future work should explore how variations on stochastic approximation algorithms will produce new non-Bayesian update rules for more general problems. Furthermore, we have modeled interactions between agents as exchanges of local probability distributions (i.e., beliefs) between neighboring nodes in a graph. It remains open to understanding to what extent this can be reduced when agents transmit only an approximate summary of their beliefs. We anticipate that future work will also consider the effect of parametric approximations allowing nodes to communicate only a finite number of parameters coming from Gaussian Mixture Models or Particle Filters.

Acknowledgment

We would like to acknowledge support for this project from the National Science Foundation under grant no. CPS 15-44953 and by the Office of Naval Research under grant no. N00014-17-1-2195.

Appendix

Proof of Proposition 1

Proof: We need to show that the density $d\mu^i_{k+1}$ associated with the probability measure μ^i_{k+1} defined by (7) minimizes Problem (6). To do so, let $G(\pi)$ be the objective function for Problem (6), i.e.,

$$G(\pi) = \langle -\log p^i_\theta(x^i_{k+1}), \pi\rangle + \sum_{j=1}^n a_{ij} D_{KL}(\pi\|d\mu^j_k).$$

Next, we add and subtract the KL divergence between π and the density $d\mu^i_{k+1}$ to obtain

$$\begin{aligned} G(\pi) &= \langle -\log p^i_\theta(x^i_{k+1}), \pi\rangle + \sum_{j=1}^n a_{ij} D_{KL}(\pi\|d\mu^j_k) - \\ &\quad - D_{KL}\left(\pi\|d\mu^i_{k+1}\right) + D_{KL}\left(\pi\|d\mu^i_{k+1}\right) \\ &= \langle -\log p^i_\theta(x^i_{k+1}), \pi\rangle + D_{KL}\left(\pi\|d\mu^i_{k+1}\right) + \\ &\quad + \sum_{j=1}^n a_{ij}\mathbb{E}_\pi \log \frac{d\mu^i_{k+1}}{d\mu^j_k}. \end{aligned}$$

Now, from (7) it follows that

$$\begin{aligned} G(\pi) &= \langle -\log p^i_\theta(x^i_{k+1}), \pi\rangle + D_{KL}\left(\pi\|d\mu^i_{k+1}\right) + \\ &\quad \sum_{j=1}^n a_{ij}\mathbb{E}_\pi \log\left(\frac{1}{d\mu^j_k}\frac{1}{Z^i_{k+1}}\prod_{l=1}^n \left(d\mu^l_k\right)^{a_{il}} p^i_\theta(x^i_{k+1})\right) \\ &= \langle -\log p^i_\theta(x^i_{k+1}), \pi\rangle + D_{KL}\left(\pi\|d\mu^i_{k+1}\right) \\ &\quad - \log Z^i_{k+1} + \langle \log p^i_\theta(x^i_{k+1}), \pi\rangle \\ &\quad + \sum_{j=1}^n a_{ij}\mathbb{E}_\pi \log\left(\frac{1}{d\mu^j_k}\prod_{l=1}^n \left(d\mu^l_k\right)^{a_{il}}\right) \\ &= -\log Z^i_{k+1} + D_{KL}\left(\pi\|d\mu^i_{k+1}\right) - \sum_{j=1}^n a_{ij}\mathbb{E}_\pi \log d\mu^j_k \\ &\quad + \sum_{l=1}^n a_{il}\mathbb{E}_\pi \log d\mu^l_k \\ &= -\log Z^i_{k+1} + D_{KL}\left(\pi\|d\mu^i_{k+1}\right), \end{aligned} \tag{15}$$

where $Z^i_{k+1} = \int_\theta p^i_\theta(x^i_{k+1})\prod_{j=1}^n (d\mu^j_k(\theta))^{a_{ij}}$ is the corresponding normalizing constant.

The first term in (15) does not depend on the distribution π. Thus, we conclude that the solution to (6) is the density $\pi^* = d\mu^i_{k+1}$ as defined in (7) (almost everywhere). ■

Proof of Lemma 4

Proof: Initially, note that

$$\begin{aligned} &\mathbb{P}\left[\sum_{t=1}^k\sum_{j=1}^n [A^{k-t}]_{ij}\log\frac{dQ^j}{dP^j}(X^j_t) \geq y\right] \\ &= \mathbb{P}\left[\log\left(\prod_{t=1}^k\prod_{j=1}^n \left(\frac{dQ^j}{dP^j}(X^j_t)\right)^{[A^{k-t}]_{ij}}\right) \geq y\right] \\ &= \mathbb{P}\left[\sqrt{\prod_{t=1}^k\prod_{j=1}^n \left(\frac{dQ^j}{dP^j}(X^j_t)\right)^{[A^{k-t}]_{ij}}} \geq \exp\left(\frac{y}{2}\right)\right]. \end{aligned}$$

By the Markov inequality we have

$$\mathbb{P}\left[\sum_{t=1}^k\sum_{j=1}^n [A^{k-t}]_{ij}\log\frac{dQ^j}{dP^j}(X^j_t) \geq y\right]$$

$$\leq \exp\left(-\frac{y}{2}\right)\mathbb{E}\left[\prod_{t=1}^{k}\prod_{j=1}^{n}\sqrt{\left(\frac{dQ^j}{dP^j}(X_t^j)\right)^{[A^{k-t}]_{ij}}}\right]$$
$$\leq \exp\left(-\frac{y}{2}\right)\prod_{t=1}^{k}\prod_{j=1}^{n}\mathbb{E}\left[\sqrt{\left(\frac{dQ^j}{dP^j}(X_t^j)\right)}\right]^{[A^{k-t}]_{ij}}$$
$$= \exp\left(-\frac{y}{2}\right)\prod_{t=1}^{k}\prod_{j=1}^{n}\rho(Q^j,P^j)^{[A^{k-t}]_{ij}},$$

where the last inequality follows from the definition of the Hellinger affinity function $\rho(Q,P)$ and Jensen's inequality.

Moreover, it follow from $\rho(Q^j,P^j)=1-h^2(Q^j,P^j)$ and $1-x\leq\exp(-x)$ for $x\in[0,1]$ that

$$\prod_{t=1}^{k}\prod_{j=1}^{n}\rho(Q^j,P^j)^{[A^{k-t}]_{ij}} \leq \exp\left(-\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij}h^2(Q^j,P^j)\right). \quad (16)$$

Now, by adding and subtracting $\sum_{t=1}^{k}\frac{1}{n}\sum_{j=1}^{n}h^2(Q^j,P^j)$ we have

$$\mathbb{P}\left[\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij}\log\frac{dQ^j}{dP^j}(X_t^j)\geq y\right]$$
$$\leq \exp\left(-\frac{y}{2}-\sum_{t=1}^{k}\sum_{j=1}^{n}\left([A^{k-t}]_{ij}-\frac{1}{n}\right)h^2(Q^j,P^j) -\frac{k}{n}\sum_{j=1}^{n}h^2(Q^j,P^j)\right)$$
$$\leq \exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{n}\sum_{j=1}^{n}h^2(Q^j,P^j)\right).$$

Finally, the last line above follows from Lemma 3 applied to the second term inside the exponential. ∎

PROOF OF THEOREM 5

Proof: We are going to focus on bounding the beliefs of a measurable set B, such that $\theta^*\in B$. For such a set, it follows by induction from (14) that

$$\mu_k^i(B)=\frac{1}{Z_k^i}\sum_{\theta\in B}\prod_{j=1}^{n}\mu_0^j(\theta)^{\left[A^k\right]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}p_\theta^j(X_t^j)^{\left[A^{k-t}\right]_{ij}}$$
$$=\left(1+\frac{\sum_{\theta\in B^c}\prod_{j=1}^{n}\mu_0^j(\theta)^{\left[A^k\right]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}p_\theta^j(X_t^j)^{\left[A^{k-t}\right]_{ij}}}{\sum_{\theta\in B}\prod_{j=1}^{n}\mu_0^j(\theta)^{[A^k]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}}}\right)^{-1}$$
$$\geq 1-\frac{\sum_{\theta\in B^c}\prod_{j=1}^{n}\mu_0^j(\theta)^{\left[A^k\right]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}p_\theta^j(X_t^j)^{\left[A^{k-t}\right]_{ij}}}{\sum_{\theta\in B}\prod_{j=1}^{n}\mu_0^j(\theta)^{[A^k]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}p_\theta^j(X_t^j)^{[A^{k-t}]_{ij}}},$$

where Z_k^i is the appropriate normalization constant. Moreover, for $\theta^*\in B$ it follows that

$$\mu_k^i(B)\geq 1-\sum_{\theta\in B^c}\prod_{j=1}^{n}\left(\frac{\mu_0^j(\theta)}{\mu_0^j(\theta^*)}\right)^{\left[A^k\right]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}\left(\frac{p_\theta^j(X_t^j)}{p^j(X_t^j)}\right)^{\left[A^{k-t}\right]_{ij}}, \quad (17)$$

The relation in (17) describes the iterative averaging of products of density functions, for which we can use Lemma 4 with $Q=P_\theta$ and $P=P_{\theta^*}$. Then,

$$\mathbb{P}\left[\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij}\log\frac{p_\theta^j(X_t^j)}{p^j(X_t^j)}\geq y\right] \leq \exp\left(-\frac{y}{2}+\frac{4\log n}{1-\delta}-\frac{k}{n}\sum_{j=1}^{n}h^2(P_\theta^j,P^j)\right).$$

We will set $y=\frac{1}{n}\sum_{j=1}^{n}\log\left(\omega\mu_0^j(\theta^*)\right)$, which is equivalent to $y=\log(\omega\mu_0(\theta^*))$ under the assumption that initial beliefs are identical among agents. For simplicity of notation, we will remove the agent super index in the initial beliefs, i.e., $\mu_0(\theta)=\mu_0^j(\theta)$ for $j\in[1,\cdots,n]$ and $\theta\in\Theta$. Thus, by defining the set

$$\Gamma_B=\left\{\boldsymbol{X}_k\mid\sup_{\theta\in B^c}\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij}\log\frac{p_\theta^j(X_t^j)}{p^j(X_t^j)}\geq\log(\omega\mu_0(\theta^*))\right\},$$

where $\boldsymbol{X}_k=\{X_t^j;t=1,\ldots,k,j=1,\cdots,n\}$, we obtain

$$\mathbb{P}\left[\Gamma_B\right]\leq\exp\left(\frac{4\log n}{1-\delta}-\frac{\log\left(\omega\mu_0(\theta^*)\right)}{2}\right)\times \times\sum_{\theta\in B^c}\exp\left(-\frac{k}{n}\sum_{j=1}^{n}h^2(P_\theta^j,P^j)\right).$$

Which in turn implies

$$\sqrt{\omega\mu_0(\theta^*)}\mathbb{P}\left[\Gamma_B\right] \leq\exp\left(\frac{4\log n}{1-\delta}\right)\sum_{\theta\in B^c}\exp\left(-\frac{k}{n}\sum_{j=1}^{n}h^2(P_\theta^j,P^j)\right).$$

Now, we let the set B be the Hellinger ball of a radius r centered at θ^* and define a cover (as described in Section V-A) to exploit the representation of $\mathcal{B}_r^c$ as the union of concentric Hellinger annuli, for which we have

$$\sqrt{\omega\mu_0(\theta^*)}\mathbb{P}\left[\Gamma_{\mathcal{B}_r^c}\right] \leq\exp\left(\frac{4\log n}{1-\delta}\right)\sum_{l=1}^{L-1}\sum_{\theta\in\mathcal{F}_l}\exp\left(-\frac{k}{n}\sum_{j=1}^{n}h^2(P_\theta^j,P^j)\right)$$
$$\leq\exp\left(\frac{4\log n}{1-\delta}\right)\sum_{l=1}^{L-1}\mathcal{N}_{r_l}\exp\left(-kr_{l+1}^2\right).$$

We seek a k large enough that the above probability is below σ. Thus, let us define the value of N as

$$N=\inf\left\{t\geq1\middle|\exp\left(\frac{4\log n}{1-\delta}\right)\sum_{l=1}^{L-1}\mathcal{N}_{r_l}\exp\left(-tr_{l+1}^2\right) <\sigma\sqrt{\omega\mu_0(\theta^*)}\right\}.$$

Thus, for all $k \geq N$ with probability $1-\sigma$, for all $\theta \in \mathcal{B}_r^c$

$$\sum_{t=1}^{k}\sum_{j=1}^{n}[A^{k-t}]_{ij}\log\frac{p_\theta^j(X_t^j)}{p^j(X_t^j)} \leq \log\left(\omega\mu_0(\theta^*)\right).$$

Thus, from (17) with probability $1-\sigma$ we have:

$$\begin{aligned}
\mu_k^i\left(\mathcal{B}_r^c\right) &\geq 1-\sum_{\theta\in B^c}\prod_{j=1}^{n}\left(\frac{\mu_0^j(\theta)}{\mu_0^j(\theta^*)}\right)^{\left[A^k\right]_{ij}}\prod_{t=1}^{k}\prod_{j=1}^{n}\left(\frac{p_\theta^j(X_t^j)}{p^j(X_t^j)}\right)^{\left[A^{k-t}\right]_{ij}} \\
&\geq 1-\sum_{\theta\in\mathcal{B}_r^c}\prod_{j=1}^{n}\left(\frac{\mu_0^j(\theta)}{\mu_0^j(\theta^*)}\right)^{\left[A^k\right]_{ij}}\exp\left(\log\left(\omega\mu_0(\theta^*)\right)\right) \\
&\geq 1-\sum_{\theta\in\mathcal{B}_r^c}\frac{\mu_0(\theta)}{\mu_0(\theta^*)}\exp\left(\log\left(\omega\mu_0(\theta^*)\right)\right) \\
&\geq 1-\sum_{\theta\in\mathcal{B}_r^c}\frac{\mu_0(\theta)}{\mu_0(\theta^*)}\mu_0(\theta^*)\omega \geq 1-\mu_0(\mathcal{B}_r^c)\omega \geq 1-\omega.
\end{aligned}$$

This completes the proof. ∎

REFERENCES

[1] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[2] K. Rahnama Rad and A. Tahbaz-Salehi, "Distributed parameter estimation in networks," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 5050–5055, 2010.

[3] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron, "Distributed bayesian hypothesis testing in sensor networks," in *Proceedings of the American Control Conference*, pp. 5369–5374, 2004.

[4] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Networked Embedded Sensing and Control*, pp. 169–182, Springer, 2006.

[5] R. J. Aumann, "Agreeing to disagree," *The Annals of Statistics*, vol. 4, no. 6, pp. 1236–1239, 1976.

[6] V. Borkar and P. P. Varaiya, "Asymptotic agreement in distributed estimation," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 650–655, 1982.

[7] J. N. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Transactions on Automatic Control*, vol. 29, no. 1, pp. 42–50, 1984.

[8] C. Genest, J. V. Zidek, *et al.*, "Combining probability distributions: A critique and an annotated bibliography," *Statistical Science*, vol. 1, no. 1, pp. 114–135, 1986.

[9] R. Cooke, "Statistics in expert resolution: A theory of weights for combining expert opinion," in *Statistics in Science* (R. Cooke and D. Costantini, eds.), vol. 122 of *Boston Studies in the Philosophy of Science*, pp. 41–72, Springer Netherlands, 1990.

[10] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[11] G. L. Gilardoni and M. K. Clayton, "On reaching a consensus using degroot's iterative pooling," *The Annals of Statistics*, vol. 21, no. 1, pp. 391–401, 1993.

[12] J. A. Gubner, "Distributed estimation and quantization," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1456–1459, 1993.

[13] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1631–1639, 2005.

[14] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[15] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.

[16] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.

[17] E. Mossel and O. Tamuz, "Efficient bayesian learning in social networks with gaussian estimators," *arXiv preprint arXiv:1002.0747*, 2010.

[18] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[19] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Business School Research Paper*, no. 13-28, 2013.

[20] S. Shahrampour and A. Jadbabaie, "Exponentially fast parameter estimation in networks using distributed dual averaging," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 6196–6201, 2013.

[21] B. Golub and M. O. Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, pp. 112–149, 2010.

[22] D. Acemoglu, A. Nedić, and A. Ozdaglar, "Convergence of rule-of-thumb learning rules in social networks," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 1714–1720, 2008.

[23] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[24] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[25] A. Olshevsky, "Linear time average consensus on fixed graphs and implications for decentralized optimization and multi-agent control," *preprint arXiv:1411.4186*, 2014.

[26] E. Mossel, A. Sly, and O. Tamuz, "Asymptotic learning on bayesian social networks," *Probability Theory and Related Fields*, vol. 158, no. 1-2, pp. 127–157, 2014.

[27] A. Lalitha, T. Javidi, and A. D. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6161–6179, 2018.

[28] L. Qipeng, Z. Jiuhua, and W. Xiaofan, "Distributed detection via bayesian updates and consensus," in *34th Chinese Control Conference (CCC)*, pp. 6992–6997, 2015.

[29] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," *IEEE Transactions on Automatic Control*, vol. 61, pp. 3256–3268, Nov 2016.

[30] S. Shahrampour, M. Rahimian, and A. Jadbabaie, "Switching to learn," in *Proceedings of the American Control Conference*, pp. 2918–2923, 2015.

[31] M. A. Rahimian, S. Shahrampour, and A. Jadbabaie, "Learning without recall by random walks on directed graphs," *preprint arXiv:1509.04332*, 2015.

[32] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed non-bayesian learning," *preprint arXiv:1508.05161*, 2015.

[33] A. Nedić, A. Olshevsky, and C. A. Uribe, "Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs," in *Proceedings of the American Control Conference*, pp. 5884–5889, 2015.

[34] A. Nedić, A. Olshevsky, and C. A. Uribe, "Network independent rates in distributed learning," in *Proceedings of the American Control Conference*, pp. 1072–1077, 2016.

[35] L. Su and N. H. Vaidya, "Asynchronous distributed hypothesis testing in the presence of crash failures," *University of Illinois at Urbana-Champaign, Tech. Rep*, 2016.

[36] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, "A theory of non-Bayesian social learning," *Econometrica*, vol. 86, no. 2, pp. 445–490, 2018.

[37] A. Mitra, J. A. Richards, and S. Sundaram, "A new approach for distributed hypothesis testing with extensions to Byzantine-resilience," in *American Control Conference (ACC)*, pp. 261–266, IEEE, 2019.

[38] A. Mitra, J. A. Richards, and S. Sundaram, "A communication-efficient algorithm for exponentially fast non-Bayesian learning in networks," in *IEEE 58th Conference on Decision and Control (CDC)*, pp. 8347–8352, IEEE, 2019.

[39] J. Z. Hare, C. A. Uribe, L. Kaplan, and A. Jadbabaie, "Non-bayesian social learning with uncertain models," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4178–4193, 2020.

[40] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Distributed detection and estimation in wireless sensor networks," *preprint arXiv:1307.1448*, 2013.

[41] A. Nedić, A. Olshevsky, and C. A. Uribe, "A tutorial on distributed (non-bayesian) learning: Problem, algorithms and results," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6795–6801, Dec 2016.

[42] L. Birgé, "About the non-asymptotic behaviour of bayes estimators," *Journal of Statistical Planning and Inference*, vol. 166, pp. 67–77, 2015.

[43] C. A. Uribe, A. Olshevsky, and A. Nedić, "Non-asymptotic concentration rates in cooperative learning part ii: Inference on compact hypothesis sets," *Submitted*, 2020.

[44] A. Nedić, A. Olshevsky, and C. A. Uribe, "Distributed learning with infinitely many hypotheses," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6321–6326, Dec 2016.

[45] S. Ghosal, "A review of consistency and convergence of posterior distribution," in *Varanashi Symposium in Bayesian Inference, Banaras Hindu University*, 1997.

[46] L. Schwartz, "On bayes procedures," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 4, no. 1, pp. 10–26, 1965.

[47] S. Ghosal, J. K. Ghosh, and A. W. Van Der Vaart, "Convergence rates of posterior distributions," *Annals of Statistics*, pp. 500–531, 2000.

[48] S. Ghosal, A. Van Der Vaart, *et al.*, "Convergence rates of posterior distributions for noniid observations," *The Annals of Statistics*, vol. 35, no. 1, pp. 192–223, 2007.

[49] V. Rivoirard, J. Rousseau, *et al.*, "Posterior concentration rates for infinite dimensional exponential families," *Bayesian Analysis*, vol. 7, no. 2, pp. 311–334, 2012.

[50] M. Rabbat, R. Nowak, and J. Bucklew, "Robust decentralized source localization via averaging," in *IEEE International Conference on Acoustics, Speech, and Signal Processing.*, vol. 5, pp. 1057–1060, 2005.

[51] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.

[52] A. Nedić and S. Lee, "On stochastic subgradient mirror-descent algorithm with weighted averaging," *SIAM Journal on Optimization*, vol. 24, no. 1, pp. 84–107, 2014.

[53] B. Dai, N. He, H. Dai, and L. Song, "Scalable bayesian inference via particle mirror descent," *preprint arXiv:1506.03101*, 2015.

[54] M. Rabbat, "Multi-agent mirror descent for decentralized stochastic optimization," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*, pp. 517–520, IEEE, 2015.

[55] A. Zellner, "Optimal information processing and bayes's theorem," *The American Statistician*, vol. 42, no. 4, pp. 278–280, 1988.

[56] S. G. Walker, "Bayesian inference via a minimization rule," *Sankhyā: The Indian Journal of Statistics (2003-2007)*, vol. 68, no. 4, pp. 542–553, 2006.

[57] T. P. Hill and M. Dall'Aglio, "Bayesian posteriors without bayes' theorem," *preprint arXiv:1203.0251*, 2012.

[58] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.

[59] A. Juditsky, P. Rigollet, A. B. Tsybakov, *et al.*, "Learning by mirror averaging," *The Annals of Statistics*, vol. 36, no. 5, pp. 2183–2206, 2008.

[60] G. Lan, A. Nemirovski, and A. Shapiro, "Validation analysis of mirror descent stochastic approximation method," *Mathematical programming*, vol. 134, no. 2, pp. 425–458, 2012.

[61] B. A. Frigyik, S. Srivastava, and M. R. Gupta, "Functional bregman divergence and bayesian estimation of distributions," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5130–5139, 2008.

[62] J. Li, G. Li, Z. Wu, and C. Wu, "Stochastic mirror descent method for distributed multi-agent optimization," *Optimization Letters*, pp. 1–19, 2016.

[63] C. W. Fox and S. J. Roberts, "A tutorial on variational bayesian inference," *Artificial intelligence review*, vol. 38, no. 2, pp. 85–95, 2012.

[64] M. J. Beal, *Variational algorithms for approximate Bayesian inference*. University of London United Kingdom, 2003.

[65] B. Dai, N. He, H. Dai, and L. Song, "Provable bayesian inference via particle mirror descent," in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pp. 985–994, 2016.

[66] B. O. Koopman, "On distributions admitting a sufficient statistic," *Transactions of the American Mathematical society*, vol. 39, no. 3, pp. 399–409, 1936.

[67] G. Darmois, "Sur les lois de probabilitéa estimation exhaustive," *CR Acad. Sci. Paris*, vol. 260, no. 1265, p. 85, 1935.

[68] D. R. Clark and C. A. Thayer, "A primer on the exponential family of distributions," in *Casualty Actuarial Society Spring Forum*, pp. 117–148, Citeseer, 2004.

[69] M. Kupperman, "Probabilities of hypotheses and information-statistics in sampling from exponential-class populations," *Selected Mathematical Papers*, vol. 29, no. 2, p. 57, 1964.

[70] E. B. Andersen, "Sufficiency and exponential families for discrete sample spaces," *Journal of the American Statistical Association*, vol. 65, no. 331, pp. 1248–1255, 1970.

[71] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*, vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.

[72] A. Nedić, A. Olshevsky, and C. A. Uribe, "Distributed gaussian learning over time-varying directed graphs," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 1710–1714, Nov 2016.

[73] C. Wang and B. Chazelle, "Gaussian learning-without-recall in a dynamic social network," *arXiv preprint arXiv:1609.05990*, 2016.

[74] L. LeCam, "Convergence of estimates under dimensionality restrictions," *The Annals of Statistics*, pp. 38–53, 1973.

[75] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st ed., 2008.

César A. Uribe received the M.Sc. degrees in systems and control from the Delft University of Technology, Delft, The Netherlands, and in applied mathematics from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2013 and 2016, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, in 2018.

He is currently the Louis Owen Jr. Chair and Assistant Professor with the Department of Electrical and Computer Engineering Department at Rice University, Houston, TX, USA. From 2018 to 2020, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include distributed learning and optimization, decentralized control, algorithm analysis, and computational optimal transport

Alex Olshevsky received the B.S. degrees in applied mathematics and electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, both in 2004, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2006 and 2010, respectively.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA. His research interests include control systems, optimization, and network science. Dr. Olshevsky received the National Science Foundation CAREER Award, the Air Force Young Investigator Award, the ICS Prize from INFORMS for best paper on the interface of operations research and computer science, and the SIAM paper prize for annual paper from the SIAM Journal on Control and Optimization chosen to be reprinted in SIAM Review.

Angelia Nedić (Member, IEEE) received the Ph.D. degree in computational mathematics and mathematical physics from Moscow State University, Moscow, Russia, in 1994, and the Ph.D. degree in electrical and computer science engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002.

She has worked as a Senior Engineer with BAE Systems North America, Arlington, VA, USA, and the Advanced Information Technology Division, Burlington, MA, USA. She has been a Willard Scholar Faculty Member with the University of Illinois at Urbana–Champaign, Champaign, IL, USA.

She is currently a Faculty Member with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA. Her general research interests include optimization, large-scale complex systems dynamics, variational inequalities, and games. Dr. Nedić was a recipient (jointly with her coauthors) of the best paper awards at the Winter Simulation Conference 2013 and the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) 2015.