

Joint Sensor Node Selection and State Estimation for Nonlinear Networks and Systems

Aleksandar Haber¹

Abstract

State estimation and sensor selection problems for nonlinear networks and systems are ubiquitous problems that are important for the control, monitoring, analysis, and prediction of a large number of engineered and physical systems. Sensor selection problems are extensively studied for linear networks. However, less attention has been dedicated to networks with nonlinear dynamics. Furthermore, widely used sensor selection methods relying on structural (graph-based) observability approaches might produce far from optimal results when applied to nonlinear network dynamics. In addition, state estimation and sensor selection problems are often treated separately, and this might decrease the overall estimation performance. To address these challenges, we develop a novel methodology for selecting sensor nodes for networks with nonlinear dynamics. Our main idea is to incorporate the sensor selection problem into an initial state estimation problem. The resulting mixed-integer nonlinear optimization problem is approximately solved using three methods. The good numerical performance of our approach is demonstrated by testing the algorithms on prototypical Duffing oscillator, associative memory, and chemical reaction networks. The developed codes are available online.

Index Terms

state and parameter estimation, observability, sensor selection, nonlinear systems, complex networks

I. INTRODUCTION

In a large variety of engineering and scientific fields, we are often faced with the problem of estimating states of networks with nonlinear dynamics. For example, this problem is crucial for identification, estimation, monitoring, and control of power systems, communication, traffic, biochemical, biophysical, combustion reaction, and ecological networks, as well as for other systems with nonlinear dynamics [1]–[11].

Generally speaking, state estimation for nonlinear networks consists of two steps. In the first step, which is often referred to as the *sensor node selection step*, we are interested in selecting a subset of network nodes whose states or output variables should be observed such that from this limited information we can accurately reconstruct the global network state. In the second step, which is often referred to as the *observer design step* or *state reconstruction step*, we are interested in designing an algorithm (observer) to reconstruct the global network state using the information collected from the sensor nodes assigned in the first step. The estimated states might be used for system monitoring, prediction, or control. Control node selection and control action design are dual problems to the estimation problems, for more details see [9], [12]–[15].

There is a large body of literature on sensor/control node selection for network dynamics. A detailed literature survey of all methods and approaches goes well beyond the scope and length limits of this manuscript. Consequently, in the sequel, we mention the main lines of research and recent contributions. Widely used and arguably most popular approaches for sensor (control) node

¹Department of Engineering and Environmental Science, The City University of New York, College of Staten Island, 2800 Victory Blvd, New York, NY 10314, USA. Email: aleksandar.haber@csi.cuny.edu.

selection are relying on graph-based methods stemming from structured control theory [16]–[18]. These approaches have been used in [3], [4] and in a large number of follow-up contributions, to devise methods for control and sensor node selection for complex networks. For example, the main idea of the approach proposed in [4], is to select sensor nodes by searching for strongly connected components in a graph describing node connections. This is a computationally inexpensive approach that can produce initial sensor node locations that can be used to initialize optimization-based sensor selection algorithms [9]. However, several authors have investigated and highlighted some of the drawbacks of graph-based approaches [12]–[15]. Namely, in some cases, the graph-based approaches might produce far from optimal solutions, or selected sensor nodes might lead to numerically ill-conditioned estimation problems [9]. Recent structural controllability and observability approaches, as well as recent surveys on these topics, can be found in [19]–[23].

Another line of research is to select sensor (control) nodes for linear networks by optimizing suitable observability (controllability) performance criteria depending on system Gramians, see for example [12], [13] and follow-up approaches. Such methods are mainly designed for linear network dynamics and might not be applicable to the nonlinear case. Empirical Gramians [24] of nonlinear systems have been used in [25], [26] to select actuators (control nodes) and sensor nodes for nonlinear networks and systems. The main limitation of these approaches is that the computation of empirical Gramians is a computationally challenging task even for small-sized networks. In [9], we have developed a sensor selection approach for nonlinear networks with smooth dynamics. However, the main prerequisite for applying this method and empirical Gramian approaches is *a priori* information on the initial state for which observability/controllability metrics are defined. In practice, we can rarely obtain accurate *a priori* information on the network initial state. An approach for sensor selection of nonlinear networks has been presented in [27]. This method assumes a specific form of nonlinear dynamics and its applicability to a broader class of nonlinear networks has to be further investigated. In [14], [28], [29], approaches have been developed to tackle some of the drawbacks of pure graph-based observability methods. These methods rely on a symbolic network Jacobian matrix and on symbolic observability coefficients. In principle, these methods can be coupled with nonlinear estimators to devise a complete solution to the network estimation problem. However, this idea needs to be numerically validated. Sensor and actuator placement problems for linear systems as well as for systems described by partial differential equations have been considered in [30]–[34].

To the best of our knowledge, sensor selection and state estimation problems are usually treated separately. The widely used approach is to first select sensor nodes by optimizing an observability criterion and then to design an estimator. Our numerical experience and results show that for certain classes of nonlinear networks this approach might lead to far from optimal estimation performance. *Our hypothesis is that by treating the sensor node selection and state estimation problems as a joint problem, we can improve the overall estimation performance.*

In this manuscript, we propose a novel problem formulation and novel methods for joint state estimation and sensor selection for nonlinear network dynamics. Our main idea is to incorporate the sensor selection problem into an initial state estimation problem. The resulting Mixed-Integer Nonlinear Optimization (MINO) problem is approximately solved using three methods. The first method uses the Mesh Adaptive Direct Search (MADS) algorithm [35] to solve the MINO problem. This is a derivative-free optimization method requiring only a procedure to compute the cost function. In an initial phase, the second and third methods solve a relaxed MINO problem obtained by replacing the integer constraints with constraints involving continuous variables. Then, the second and third methods use the solution of the relaxed problem to formulate and solve Mixed-Integer Linear Programming (MILP) problems. These MILP problems are solved by using a hybrid branch and bound method. The second and third methods are partly inspired by approaches for solving mixed-integer optimization problems arising in optimal predictive control [36]–[38].

The key contribution of this article is a novel unified framework for joint sensor node selection and initial state estimation. Furthermore, we devise and thoroughly numerically test three approaches for approximating the solutions of sensor selection problems for nonlinear systems. The methods developed in this paper approximately solve an objectively difficult optimization problem. Namely, in its full generality, the considered MINO problem is constrained, NP-hard, and non-convex. An additional layer of complexity to this problem comes from the fact that the optimization problem is constrained by nonlinear dynamics. Furthermore, all three developed methods aim at computing an approximate solution of the MINO problem by relaxing the integer constraints or by direct search. Due to all these layers of complexity, it is challenging to derive a generally-applicable convergence analysis of the developed methods. Thorough theoretical analysis might be possible only for relatively simple forms of nonlinear dynamics. To provide valuable insights into the behavior of the developed approaches, we investigate the performance of the methods by performing extensive simulation experiments. The good numerical performance of our methods is demonstrated by testing them on Duffing oscillator networks, associative memory networks, and on two models of chemical reaction networks. We compare our methods with exhaustive search and random selection of sensor nodes. Compared to other methods for solving MINO problems [39], such as methods based on branch and bound algorithms, our approach is relatively easy to implement, does not rely on convexification procedures that are usually case dependent, and is applicable to a broad class of smooth nonlinear systems. The developed codes are available online [40].

This manuscript is organized as follows. In Section II, we present the problem formulation. In Section III, we present the sensor selection algorithms. In Section IV we derive expressions for the derivatives of cost functions used in this paper and present some generalizations of the developed approaches. Numerical results and conclusions are presented in Section V and Section VI, respectively.

II. PROBLEM FORMULATION

In this section, we incorporate the sensor selection problem into an open-loop initial state estimation problem. This enables us to formulate the sensor selection and initial state estimation problems as a single MINO problem. Three algorithms for approximating the solution of this problem are stated in Section III.

First, we explain the used notation. Let $\mathbf{w}_k \in \mathbb{R}^n$ denote an arbitrary n -dimensional discrete-time vector at a discrete-time instant $k \in \mathbb{Z}_0^+$. The notation $\mathbf{w}_{0:S}$ denotes a lifted column vector $\mathbf{w}_{0:S} := \text{col}(\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_S)$, $\mathbf{w}_{0:S} \in \mathbb{R}^{(S+1)n}$, where $S \in \mathbb{Z}_0^+$. The notation $\text{col}(\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_S)$ denotes an operator defined by $\text{col}(\cdot) = [\mathbf{w}_0^T, \mathbf{w}_1^T, \dots, \mathbf{w}_S^T]^T$. The notation $\mathbf{g}[\mathbf{w}]$ denotes a vector function of the vector argument \mathbf{w} . The notation $\text{diag}(\mathbf{w}) \in \mathbb{R}^{n \times n}$ denotes a diagonal matrix with the entries of the vector \mathbf{w} on its main diagonal. The notation I_S denotes an $S \times S$ identity matrix.

We consider a nonlinear state-space model

$$\dot{\mathbf{x}} = \mathbf{f}[\mathbf{x}], \quad (1)$$

$$\mathbf{y} = C\mathbf{x}, \quad (2)$$

where $\mathbf{x} \in \mathbb{R}^{Nn}$ is a global network state, $\mathbf{f}[\cdot] : \mathbb{R}^{Nn} \rightarrow \mathbb{R}^{Nn}$ is a nonlinear vector function that is sufficiently smooth, $\mathbf{y} \in \mathbb{R}^M$ is an observed global output vector, and $C \in \mathbb{R}^{M \times Nn}$ is an output matrix. We assume that the network consists of N nodes, where a local state of each node is n -dimensional, that is, $\mathbf{x} = \text{col}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)})$, where $\mathbf{x}^{(i)} \in \mathbb{R}^n$, $i = 1, \dots, N$, is the local state of the i -th node. We assume that the output vector \mathbf{y} consists of M entries $y^{(j)} \in \mathbb{R}$, where the index j can take any value from the set $\{1, 2, \dots, N\}$. The entry $y^{(j)}$ is a local observation at a prescribed node. A node is a *sensor node* if its output is being observed. For presentation clarity, we have assumed a specific form of the output equation (2). The approaches

presented in this paper can be generalized to other nonlinear forms of the output equation. The integer M denotes the number of sensor nodes. Next, let $\mathbf{z} = \text{col}(y^{(1)}, y^{(2)}, \dots, y^{(N)})$, $\mathbf{z} \in \mathbb{R}^N$, denote an output vector obtained assuming that all network nodes are sensor nodes. We assume that the sensor node outputs are available at discrete time samples $t = kh$, where $k = 0, 1, 2, \dots$, and h is a discretization constant. Furthermore, let $\mathbf{y}_k := \mathbf{y}[kh]$, $\mathbf{z}_k := \mathbf{z}[kh]$, and $\mathbf{x}_k := \mathbf{x}[kh]$.

The locations of the sensor nodes are encoded in the structure of the matrix C . Knowing the model of the system dynamics, given by the equation (1), our goal is to determine an optimal (sparsity) structure of the matrix C , under the condition that sensors can be placed on a fixed number of nodes that is usually smaller than N . At the same time, we want to ensure that the selected matrix C will produce accurate results when used to design an estimator or an observer. We consider the sensor selection and initial state estimation problems under the following scenarios. In the first scenario, we can simulate the model of the network dynamics (1) for a user-selected initial condition. In this way, we can obtain the state trajectory $\{\mathbf{x}_k\}$, and consequently, we can construct the sequence $\{\mathbf{z}_k\}$. Then, using the sequence $\{\mathbf{z}_k\}$, and the model (1), under the condition that only M_{\max} of the sensor nodes can be used, the goal is to select the optimal structure of the matrix C such that the initial state of the real system can be accurately reconstructed. After we have determined the sensor locations, we can place sensors on predetermined locations, and these sensors can be used to collect the output sequence $\{\mathbf{y}_k\}$ from the real system that can be used for state estimation. In the second scenario, instead of simulating the system response (for a selected initial condition), we can place sensors on all nodes of the real system to obtain the sequence $\{\mathbf{z}_k\}$. Then, using this sequence collected from the real system, we can find the optimal locations of M_{\max} sensors. After completing the sensor selection procedure, all the sensors except M_{\max} selected sensors should be removed from the network, and the initial state should be estimated. This scenario is only feasible for small or medium-sized networks due to practical and economical constraints. It should be emphasized that in the first scenario, the user-selected initial condition necessary to obtain the data sequence $\{\mathbf{z}_k\}$ is usually different from the unknown initial conditions of the real physical system. There are two approaches to minimize this uncertainty. The first approach is to choose the initial conditions to be in the expected range of the initial conditions of the system. The second approach is discussed in Section IV-B.

We can formally define the joint problem of selecting the sensor nodes and estimating the initial states as follows.

Problem 1 (Joint Sensor Selection and Initial State Estimation): Given the maximal number of sensor nodes, denoted by M_{\max} , $M_{\max} \leq N$, and given the length of the observation horizon, denoted by L , determine:

- 1) Locations of sensor nodes using the sequence of data samples $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_L\}$. These data samples can be obtained by using the two approaches. In the first approach, the data samples are obtained by simulating the system (1) for the user-selected initial condition. In the second approach, the samples are collected by placing sensors on all nodes of the real system.
- 2) An estimate of the real system initial state \mathbf{x}_0 determined from the sequence of output data samples $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$ collected from the real system by the assigned sensor nodes.

To mathematically formulate the sensor selection problem, we represent the continuous-time dynamics (1) in the discrete-time domain [9], [41]–[43]. The simplest discretization approach is based on the Forward Euler (FE) method [44]:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + hf[\mathbf{x}_{k-1}], \quad (3)$$

The FE dynamics (3) approximates relatively well the continuous-time dynamics when h is small and when the dynamics is not stiff. Stiff network dynamics are characterized by time constants of local nodes that significantly differ in magnitude. A large number of systems such as reaction-diffusion systems, chemical reaction networks, and other systems coupling multi-physics

phenomena have stiff dynamics. To allow for larger values of the discretization constant and to be able to accurately represent stiff network dynamics, we employ the Trapezoidal Implicit (TI) discretization method [44]:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + 0.5h(\mathbf{f}[\mathbf{x}_k] + \mathbf{f}[\mathbf{x}_{k-1}]). \quad (4)$$

Another option is to use the implicit Runge-Kutta method [9], [44], however, for brevity, we do not use such a method in this manuscript. To solve the sensor selection method we have to simulate the system dynamics many times. The main disadvantage of the TI method over the FE method is that in every simulation step k , we need to solve the system of nonlinear equations (4) for \mathbf{x}_k . This generally results in $\mathcal{O}(N^3n^3)$ computational complexity. In sharp contrast, in every step k , the computational complexity of the FE method is $\mathcal{O}(Nn)$. This implies that the methods presented in the next section will have a larger computational complexity for the TI dynamics. For notational brevity, both FE and TI discretized dynamics are denoted by the following equation

$$\mathbf{x}_k = \mathbf{g}[\mathbf{x}_k, \mathbf{x}_{k-1}]. \quad (5)$$

Next, we introduce a parametrized output equation relating states and \mathbf{z}_k :

$$\mathbf{z}_k = C_\theta[\theta]\mathbf{x}_k, \quad (6)$$

where $\theta \in \{0, 1\}^N$ is a binary parametrization vector and $C_\theta[\theta] \in \mathbb{R}^{N \times Nn}$ is a parametrized output matrix. Depending on the structure of a system used for performing numerical experiments, we assume two parametrization forms of $C_\theta[\theta]$. For networks for which $n = 1$, we assume $C_\theta = \text{diag}(\theta)$. For $n > 1$, C_θ is a block diagonal matrix with the i -th block equal to $C_i = [\theta_i \ 0 \ \dots \ 0] \in \mathbb{R}^{1 \times n}$, where θ_i is the i -th entry of θ . Our idea for developing the sensor selection method originates from the observability definition for discrete-time systems [9], [45]. Namely, the observed output sequence $\mathbf{y}_{0:L}$ and the initial state \mathbf{x}_0 are related by a nonlinear function \mathbf{w}

$$\mathbf{y}_{0:L} = \mathbf{w}[\mathbf{x}_0], \quad (7)$$

where the notation $\mathbf{y}_{0:L}$ is used to denote a lifted vector (see the beginning of Section II for used notation) and the nonlinear function $\mathbf{w}[\mathbf{x}_0] : \mathbb{R}^{Nn} \rightarrow \mathbb{R}^{(L+1)M}$ is defined by

$$\mathbf{w}[\mathbf{x}_0] := (I_{L+1} \otimes C)\mathbf{x}_{0:L} = (I_{L+1} \otimes C)\mathbf{x}_{0:L}[\mathbf{x}_0], \quad (8)$$

where $I_{L+1} \in \mathbb{R}^{(L+1) \times (L+1)}$ is an identity matrix and \otimes denotes the Kronecker product. The state sequence $\mathbf{x}_{0:L}[\mathbf{x}_0]$ is only a function of the initial state \mathbf{x}_0 since from (5) we obtain telescopic equalities

$$\mathbf{x}_1 = \mathbf{g}[\mathbf{x}_1, \mathbf{x}_0], \quad \mathbf{x}_2 = \mathbf{g}[\mathbf{x}_2, \mathbf{x}_1], \quad \dots, \quad (9)$$

$$\mathbf{x}_L = \mathbf{g}[\mathbf{x}_L, \mathbf{x}_{L-1}]. \quad (10)$$

Consequently, for known \mathbf{x}_0 we can compute \mathbf{x}_k , $k = 1, 2, \dots, L$, by solving the nonlinear equations (for the TI dynamics) or by simply propagating the telescopic equations (for the FE dynamics). Let us recall the *uniform observability definition*. A discrete-time dynamical system is uniformly observable on a set if there exists $L > 0$ such that $\mathbf{w}[\mathbf{x}_0]$ is an injective function with respect to initial state \mathbf{x}_0 [45]. From the practical point of view, this means that the system is observable if we can uniquely solve the system of nonlinear equations (7) for \mathbf{x}_0 . This problem can be reformulated as a nonlinear optimization problem

$$\min_{\mathbf{x}_0} \|\mathbf{y}_{0:L} - \mathbf{w}[\mathbf{x}_0]\|_2^2, \quad \text{subject to (9)-(10)}. \quad (11)$$

Similarly to the continuous-time nonlinear systems, sufficient observability conditions for discrete-time dynamics can be formulated. The uniform observability condition is that the Jacobian matrix of $\mathbf{w}[\mathbf{x}_0]$ is of full-rank [9], [45]. In [9] we have

developed a sensor selection method that optimizes a criterion that involves the Jacobian matrix. However, the Jacobian matrix is defined locally for a specific initial state \mathbf{x}_0 , and consequently, the observability condition will vary from one state to another. Also, in practice the initial state knowledge is limited, and consequently, the sensor selection approach that is solely based on optimizing a criterion depending on the Jacobian matrix might produce inaccurate results. *This motivates us to incorporate the sensor selection problem into the state estimation problem (11) and to jointly solve the resulting problem for the initial state and sensor locations.* Following this idea, we formulate the joint problem of selecting the sensor nodes and estimating the initial state as the solution of the following MINO problem:

$$(\mathbf{P1}) \min_{\mathbf{x}_0, \boldsymbol{\theta}} \left\| \mathbf{z}_{0:L} - (I_{L+1} \otimes C_{\boldsymbol{\theta}}[\boldsymbol{\theta}]) \mathbf{x}_{0:L}[\mathbf{x}_0] \right\|_2^2, \quad (12a)$$

$$\text{subject to } \mathbf{x}_i = \mathbf{g}[\mathbf{x}_i, \mathbf{x}_{i-1}], \quad i = 1, 2, \dots, L, \quad (12b)$$

$$\sum_{l=1}^N \theta_l \leq M_{\max}, \quad \left(\sum_{l=1}^N \theta_l = M_{\max} \right), \quad (12c)$$

$$\theta_l \in \{0, 1\}, \quad l = 1, 2, \dots, N, \quad (12d)$$

$$\underline{\mathbf{x}}_0 \leq \mathbf{x}_0 \leq \bar{\mathbf{x}}_0. \quad (12e)$$

The constraint (12c) limits the number of selected sensor nodes. We have two options for incorporating these constraints into the optimization problem. The first option, represented by the \leq relation, ensures that the number of selected nodes is *smaller than or equal* to M_{\max} . The second option, represented by the equality relation (in the brackets), is used to ensure that the number of sensor nodes is precisely equal to M_{\max} . The second option is introduced since in our simulations we want to compare the developed algorithms with random sensor selection, and in order to ensure that the comparison is fair, we need to ensure that the selected number of sensor nodes is always constant. Also, we noticed that nonlinear solvers occasionally perform better in the case of the second option. In (12e) the notation \leq denotes the element-wise less than equal relation, $\underline{\mathbf{x}}_0 \in \mathbb{R}^{Nn}$ and $\bar{\mathbf{x}}_0 \in \mathbb{R}^{Nn}$ are lower and upper bounds on \mathbf{x}_0 . By relaxing the binary constraints and by eliminating the constraints on the maximal number of sensor nodes, from the MINO problem (12) we obtain a relaxed problem:

$$(\mathbf{P2}) \min_{\mathbf{x}_0, \boldsymbol{\theta}} \left\| \mathbf{z}_{0:L} - (I_{L+1} \otimes C_{\boldsymbol{\theta}}(\boldsymbol{\theta})) \mathbf{x}_{0:L}[\mathbf{x}_0] \right\|_2^2, \quad (13a)$$

$$\text{subject to } \mathbf{x}_i = \mathbf{g}[\mathbf{x}_i, \mathbf{x}_{i-1}], \quad i = 1, 2, \dots, L, \quad (13b)$$

$$\mathbf{0} \leq \boldsymbol{\theta} \leq \mathbf{1}, \quad (13c)$$

$$\underline{\mathbf{x}}_0 \leq \mathbf{x}_0 \leq \bar{\mathbf{x}}_0. \quad (13d)$$

In (13), $\mathbf{0} \in \mathbb{R}^N$ and $\mathbf{1} \in \mathbb{R}^N$ are the vectors of zeros and ones, respectively.

III. SENSOR SELECTION METHODS

In this section, we introduce three algorithms for approximating the solution of the MINO problem **P1** in (12). The first algorithm is based on solving the problem using the NOMAD algorithm [46]. The NOMAD algorithm is initialized with a

solution guess generated by solving the relaxed problem **P2** defined in (13). The second and third algorithms are based on computing the initial solution guess by solving the relaxed problem and then using this initial solution to formulate and solve MILPs that are defined in the sequel. The developed codes are available online [40].

A. Algorithm 1 - Mesh Adaptive Direct Search Approach

This method is summarized in Algorithm 1. The NOMAD algorithm (also known as MADS) is a derivative-free optimization method. We use a version of this algorithm implemented in the OPTI MATLAB toolbox [46]. The advantage of this approach is that it can easily be integrated with all MATLAB nonlinear solvers that are necessary to solve the system of equations (12b). Furthermore, to implement this method we only need a numerical procedure to evaluate the cost function.

In step 1 of Algorithm 1, we solve the relaxed problem **P2** defined in (13). This problem is solved using the interior point method implemented in the MATLAB function `fmincon(·)`. We use a recursive approach for solving the relaxed problem. The recursive approach is originally used for solving model predictive control problems, see Chapter 10 in [47]. However, with minor modifications, it is also applicable to our problem. The recursive approach does not consider intermediate states $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L$ in (13b) as explicit optimization variables. Instead, in every iteration, the state sequences are computed by either forward propagation of the discrete-time dynamics (3) (in the case of the FE method) or by solving the nonlinear system of equations (9)-(10) (in the case of the TI method). The nonlinear system of equations is solved using the trust-region dogleg method implemented in the MATLAB function `fsolve(·)`. The solution of the relaxed problem is used in step 2 of Algorithm 1 as an initial guess for the NOMAD solver. By generating the initial guess in this way, we significantly decrease the number of iterations of the NOMAD algorithm. To implement the NOMAD algorithm, we use the previously explained recursive approach. Once the state sequence is computed, we can evaluate the cost function (12a), and this value is given as an input to the NOMAD solver. Finally, in step 3, using the sensor nodes computed by the NOMAD algorithm, we form the matrix \hat{C} by eliminating the zero rows (corresponding to sensor nodes that are not selected) of the matrix $C_\theta[\hat{\theta}]$. For this \hat{C} , we form and solve the optimization problem (11) to compute the initial state estimate $\hat{\mathbf{x}}_0$. This problem is solved using the quasi-Newton method implemented in the MATLAB function `fminunc(·)`.

Algorithm 1 Sensor Selection Using NOMAD Method

inputs: In the first phase (steps 1-2), the inputs are the output sequence $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_L\}$ and M_{\max} . In the second phase (step 3), the input is the output sequence $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$ collected by the assigned sensor nodes.

outputs: The optimal sensor selection vector $\hat{\theta}$ and the initial state estimate $\hat{\mathbf{x}}_0$.

1. initial solution: Solve the relaxed problem **P2** in (13). Let the solution of this problem be denoted by $(\tilde{\theta}, \tilde{\mathbf{x}}_0)$.

2. solve: Using $(\tilde{\theta}, \tilde{\mathbf{x}}_0)$ as an initial guess, solve the MINO problem **P1** in (12) by using the NOMAD solver. Let the solution of this problem be denoted by $(\hat{\theta}, \tilde{\mathbf{x}}_0^{(1)})$.

3. solve: Implement the sensor locations on the physical system. Using the installed sensors, collect the output sequence $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$. Form the matrix \hat{C} by eliminating the zero rows of the matrix $C_\theta(\hat{\theta})$ in (6). Setting $C := \hat{C}$, solve (11) to compute $\hat{\mathbf{x}}_0$.

In Section IV we derive the gradients of the defined cost functions and we present some generalizations of the developed approaches.

B. Algorithm 2 - Solving Relaxed and MILP Problems

Here we present an algorithm for approximating the solution of the MINO problem **P1** in (12) that is based on solving the relaxed problem **P2** in (13) and a MILP that is defined in the sequel. Let the solution of the relaxed problem be denoted by $\tilde{\mathbf{x}}_0$. Using this solution, we define the following cost function

$$J[\boldsymbol{\theta}] = \|G[\boldsymbol{\theta}]\|_1, \quad (14)$$

$$G[\boldsymbol{\theta}] = \mathbf{z}_{0:L} - (I_{L+1} \otimes C_{\boldsymbol{\theta}}[\boldsymbol{\theta}])\mathbf{x}_{0:L}[\tilde{\mathbf{x}}_0]. \quad (15)$$

The state sequence $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_L\}$ or compactly $\mathbf{x}_{0:L}[\tilde{\mathbf{x}}_0]$ is obtained by propagating the equations (9)-(10) from $\tilde{\mathbf{x}}_0$ in the case FE dynamics or by solving these equations starting from $\tilde{\mathbf{x}}_0$ in the case of the TI dynamics. It should be noted that this cost function is similar to the cost functions in (12a) and (13a), except for the ℓ_1 norm and substituted initial state. We determine optimal sensor locations as the solution of the following optimization problem:

$$\min_{\boldsymbol{\theta}} J[\boldsymbol{\theta}], \quad (16)$$

$$\text{subject to } \sum_{l=1}^N \theta_l \leq M_{\max}, \left(\sum_{l=1}^N \theta_l = M_{\max} \right), \quad (17)$$

$$\theta_l \in \{0, 1\}, \quad l = 1, 2, \dots, N. \quad (18)$$

To solve (16)-(18), we first transform this problem into an MILP problem. By using the parametrization of the matrix $C_{\boldsymbol{\theta}}[\boldsymbol{\theta}]$ introduced in Section II, we can transform the cost function (16) into the following form

$$J[\boldsymbol{\theta}] = \sum_{j=0}^L \sum_{i=1}^N |z_j^{(i)} - \theta_i \tilde{x}_{j1}^{(i)}|, \quad (19)$$

where $z_j^{(i)}$ is the i -th entry of \mathbf{z}_j , and $\tilde{x}_{j1}^{(i)}$ is the first entry of $\tilde{\mathbf{x}}_j^{(i)}$, and this vector is the i -th entry of $\tilde{\mathbf{x}}_j$ (the index j is the discrete-time instant). By introducing the slack variable vector $\mathbf{b} \in \mathbb{R}^{N(L+1)}$, and by taking into account (19), we can transform the optimization problem (16)-(18) as follows

$$(\text{MILP1}) \quad \min_{\mathbf{b}, \boldsymbol{\theta}} \sum_{j=0}^L \sum_{i=1}^N b_j^{(i)}, \quad (20a)$$

subject to

$$z_j^{(i)} - \theta_i \tilde{x}_{j1}^{(i)} \leq b_j^{(i)}, \quad \theta_i \tilde{x}_{j1}^{(i)} - z_j^{(i)} \leq b_j^{(i)}, \quad b_j^{(i)} \geq 0, \quad (20b)$$

$$\sum_{i=1}^N \theta_i \leq M_{\max}, \left(\sum_{i=1}^N \theta_i = M_{\max} \right), \quad \theta_i \in \{0, 1\}, \quad (20c)$$

$$i = 1, \dots, N, \quad j = 0, 1, \dots, L,$$

where $b_j^{(i)}$ is the i -th entry of \mathbf{b}_j , and \mathbf{b}_j is the j -th entry of $\mathbf{b} = \text{col}(\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_L)$.

The problem (20) is a MILP problem that we solve using the MATLAB function `intlinprog(·)`. The MATLAB function `intlinprog(·)` implements a hybrid method partly based on a branch and bound method. Algorithm 2 summarizes this approach for approximating the solution of the MINLP problem.

Algorithm 2 Sensor Selection by Solving the Relaxed Problem **P2** and **MILP1**

inputs: In the first phase (steps 1-3), the inputs are the output sequence $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_L\}$ and M_{\max} . In the second phase (step 4), the input is the output sequence $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$ collected by the assigned sensor nodes.

outputs: The optimal sensor selection vector $\hat{\boldsymbol{\theta}}$ and the initial state estimate $\hat{\mathbf{x}}_0$.

1. initial solution: Solve the relaxed problem **P2** in (13). Let the solution of this problem be denoted by $(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{x}}_0)$.

2. compute state sequence: Using $\tilde{\mathbf{x}}_0$ compute the state sequence $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_L\}$ by solving the telescopic equations (9)-(10).

3. solve: Using $\{\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_L\}$ form and solve (20). Let the solution of this problem be denoted by $\hat{\boldsymbol{\theta}}$.

4. solve: Implement the sensor locations on the physical system. Using the installed sensors, collect the output sequence $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$. Form the matrix \hat{C} by eliminating the zero rows of the matrix $C_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ in (6). Setting $C := \hat{C}$, solve (11) to compute $\hat{\mathbf{x}}_0$.

C. Algorithm 3 - Solving Relaxed and Binary MILP Problem

Algorithm 2 is based on computing the state sequence once the relaxed problem is solved. Due to the specific structure of the cost function (14), we can create another cost function in which the initial state is eliminated. This cost function penalizes the difference between the solution $\tilde{\boldsymbol{\theta}}$ of the relaxed problem **P2** in (13) and the optimization variable $\boldsymbol{\theta}$. The third algorithm computes optimal sensor locations by minimizing this cost function. This approach is inspired by an approach for solving mixed-integer optimal control problems in [36]–[38]. However, our problem formulation and introduced cost function differ from the ones considered in [36]–[38].

Consider the expression G in (15). The difference between the value of G for the relaxed solution $\tilde{\boldsymbol{\theta}}$ and for the binary solution $\boldsymbol{\theta}$ that we want to determine is given by

$$G[\tilde{\boldsymbol{\theta}}] - G[\boldsymbol{\theta}] = \left(I_{L+1} \otimes (C_{\boldsymbol{\theta}}[\boldsymbol{\theta}] - C_{\tilde{\boldsymbol{\theta}}}[\tilde{\boldsymbol{\theta}}]) \right) \mathbf{x}_{0:L}[\tilde{\mathbf{x}}_0]. \quad (21)$$

Our goal is to find $\boldsymbol{\theta}$ such that an upper bound on the ℓ_1 norm of this difference is minimized while ensuring that the constraints on the total number of sensor nodes are satisfied. From (21) we have

$$\|G[\tilde{\boldsymbol{\theta}}] - G[\boldsymbol{\theta}]\|_1 \leq \|I_{L+1} \otimes (C_{\boldsymbol{\theta}}[\boldsymbol{\theta}] - C_{\tilde{\boldsymbol{\theta}}}[\tilde{\boldsymbol{\theta}}])\|_1 \|\mathbf{x}_{0:L}[\tilde{\mathbf{x}}_0]\|_1. \quad (22)$$

This upper bound can be minimized by minimizing the first term on the right-hand-side of (22). This term is minimized if $\|C_{\boldsymbol{\theta}}[\boldsymbol{\theta}] - C_{\tilde{\boldsymbol{\theta}}}[\tilde{\boldsymbol{\theta}}]\|_1$ is minimized. Recalling the definition of the ℓ_1 norm, we conclude that in order to minimize $\|C_{\boldsymbol{\theta}}[\boldsymbol{\theta}] - C_{\tilde{\boldsymbol{\theta}}}[\tilde{\boldsymbol{\theta}}]\|_1$, we need to minimize the maximum of all the column sums of matrix difference $C_{\boldsymbol{\theta}}[\boldsymbol{\theta}] - C_{\tilde{\boldsymbol{\theta}}}[\tilde{\boldsymbol{\theta}}]$. Since the matrix $C_{\boldsymbol{\theta}}[\boldsymbol{\theta}]$ has a (block) diagonal structure and due to the parametrization introduced in Section II, the problem boils down to the problem of minimizing the maximal difference between the entries of $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$. Consequently, to compute optimal sensor locations, we formulate the following optimization problem:

$$\min_{\boldsymbol{\theta}} \max_i |\theta_i - \tilde{\theta}_i|, \quad (23)$$

$$\text{subject to } \sum_{l=1}^N \theta_l \leq M_{\max}, \left(\sum_{l=1}^N \theta_l = M_{\max} \right), \quad (24)$$

$$\theta_l \in \{0, 1\}, \quad l = 1, 2, \dots, N. \quad (25)$$

By introducing the slack variable $q \in \mathbb{R}_0^+$, the last optimization problem can be reformulated as follows

$$(\text{MILP2}) \quad \min_{q, \theta} q, \quad (26a)$$

$$\text{subject to } \theta_i - \tilde{\theta}_i \leq q, \quad \tilde{\theta}_i - \theta_i \leq q, \quad i = 1, 2, \dots, N, \quad (26b)$$

$$\sum_{l=1}^N \theta_l \leq M_{\max}, \quad \left(\sum_{l=1}^N \theta_l = M_{\max} \right), \quad (26c)$$

$$\theta_l \in \{0, 1\}, \quad l = 1, 2, \dots, N. \quad (26d)$$

Similarly to the **MILP1** problem in (20), the **MILP2** problem in (26) is solved by using the hybrid method implemented in the MATLAB function `intlinprog(·)`. Algorithm 3 summarizes the third approach for selecting the sensor nodes.

Algorithm 3 Sensor Selection by Solving Relaxed Problem **P2** and **MILP2**

inputs: In the first phase (steps 1-2), the inputs are the output sequence $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_L\}$ and the maximal number of sensor nodes M_{\max} . In the second phase (step 3), the inputs is the output sequence $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$ collected by the assigned sensor nodes.

outputs: The optimal sensor selection vector $\hat{\theta}$ and the initial state estimate $\hat{\mathbf{x}}_0$.

1. initial solution: Solve the relaxed problem P2 in (13). Let the solution of this problem be denoted by $(\tilde{\theta}, \tilde{\mathbf{x}}_0^{(0)})$.

2. solve: Using $\tilde{\theta}$ solve (26). Let the solution of this problem be denoted by $\hat{\theta}$

3. solve: Implement the sensor locations on the physical system. Using the installed sensors, collect the output sequence $\{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_L\}$. Form the matrix \hat{C} by eliminating the zero rows of the matrix $C_{\theta}(\hat{\theta})$. Setting $C := \hat{C}$, solve (11) to compute $\hat{\mathbf{x}}_0$.

In the general case, the formulated MINO problem **P1** in (12) is NP-hard, non-convex, and constrained by nonlinear dynamics. On the other hand, the introduced methods are approximating the solution. Due to the many layers of complexity, theoretical investigation of the degree of the suboptimality of the developed methods is challenging. To provide crucial insights into the optimality of the developed approaches, in Section V we investigate the performance of our methods by performing extensive numerical experiments. The good numerical performance of our methods is demonstrated by testing the methods on Duffing oscillator networks that are prototypical models of a number of dynamical systems, as well as on associative memory networks representing simplified memory models and on chemical reaction networks. We compare our methods with exhaustive search and random selection of sensor nodes.

IV. DERIVATIVES AND GENERALIZATIONS

To significantly speed up the computations of the solutions of the optimization problems, in this section we derive expressions for gradients and derivatives of the cost function used in the problems (12) and (13). In addition, in Section IV-B we briefly present generalizations of the optimization problems (12) and (13) that incorporate a number of output and state sequences. In this way, we can increase the robustness and decrease the sensitivity of the sensor selection procedure with respect to uncertainties and lack of knowledge of the initial state of the system.

A. Derivatives

First, we start with the derivatives of the FE and TI dynamics in (3) and (4). Similar to the recursive computation (simulation) of state sequences, derivatives are also computed in a recursive manner. From (3) it follows that for the FE dynamics, we have

$$\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_0} = \frac{\partial \mathbf{x}_{k-1}}{\partial \mathbf{x}_0} + h \frac{\partial \mathbf{x}_{k-1}}{\partial \mathbf{x}_0} \frac{\partial \mathbf{f}[\mathbf{x}_{k-1}]}{\partial \mathbf{x}_{k-1}}. \quad (27)$$

On the other hand, from (4) it follows that for the TI dynamics, we have

$$\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_0} = \frac{\partial \mathbf{x}_{k-1}}{\partial \mathbf{x}_0} + \frac{h}{2} \left(\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_0} \frac{\partial \mathbf{f}[\mathbf{x}_k]}{\partial \mathbf{x}_k} + \frac{\partial \mathbf{x}_{k-1}}{\partial \mathbf{x}_0} \frac{\partial \mathbf{f}[\mathbf{x}_{k-1}]}{\partial \mathbf{x}_{k-1}} \right). \quad (28)$$

Under a mild assumption of invertibility of $(I_{Nn} - (h/2)\partial \mathbf{f}[\mathbf{x}_k]/\partial \mathbf{x}_k)$, from the last expression, we obtain

$$\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_0} = \frac{\partial \mathbf{x}_{k-1}}{\partial \mathbf{x}_0} \left(I_{Nn} + \frac{h}{2} \frac{\partial \mathbf{f}[\mathbf{x}_{k-1}]}{\partial \mathbf{x}_{k-1}} \right) \left(I_{Nn} - \frac{h}{2} \frac{\partial \mathbf{f}[\mathbf{x}_k]}{\partial \mathbf{x}_k} \right)^{-1}. \quad (29)$$

By initializing (27) and (29) with $\partial \mathbf{x}_0/\partial \mathbf{x}_0 = I_{Nn}$, we can recursively compute the derivatives of state sequences. These derivatives will be used to compute the gradients of the cost function

$$\mathcal{W}[\boldsymbol{\theta}, \mathbf{x}_0] = \|\mathbf{z}_{0:L} - K[\boldsymbol{\theta}] \mathbf{x}_{0:L}[\mathbf{x}_0]\|_2^2, \quad K[\boldsymbol{\theta}] = (I_{L+1} \otimes C_{\boldsymbol{\theta}}[\boldsymbol{\theta}]), \quad (30)$$

that appear in the MINO problem **P1** defined in (12) and in the relaxed problem **P2** defined in (13). The gradient of the cost function is defined by

$$\nabla \mathcal{W} = \begin{bmatrix} \nabla_{\mathbf{x}_0} \mathcal{W} \\ \nabla_{\boldsymbol{\theta}} \mathcal{W} \end{bmatrix}. \quad (31)$$

For brevity we only give the final gradient expressions. We obtained

$$\nabla_{\mathbf{x}_0} \mathcal{W} = -2 \frac{\partial \mathbf{x}_{0:L}[\mathbf{x}_0]}{\partial \mathbf{x}_0} K^T[\boldsymbol{\theta}] (\mathbf{z}_{0:L} - K[\boldsymbol{\theta}] \mathbf{x}_{0:L}[\mathbf{x}_0]), \quad (32)$$

$$\frac{\partial \mathbf{x}_{0:L}[\mathbf{x}_0]}{\partial \mathbf{x}_0} = \begin{bmatrix} I_{Nn} & \frac{\partial \mathbf{x}_1}{\partial \mathbf{x}_0} & \cdots & \frac{\partial \mathbf{x}_L}{\partial \mathbf{x}_0} \end{bmatrix}, \quad (33)$$

$$\nabla_{\boldsymbol{\theta}} \mathcal{W} = -2 \begin{bmatrix} D_0 & D_1 & \cdots & D_L \end{bmatrix} (\mathbf{z}_{0:L} - K[\boldsymbol{\theta}] \mathbf{x}_{0:L}[\mathbf{x}_0]), \quad (34)$$

where $D_i = \text{diag}(C_1 \mathbf{x}_i)$ and the matrix C_1 depends on the parametrization of $C_{\boldsymbol{\theta}}[\boldsymbol{\theta}]$. In the case when $C_{\boldsymbol{\theta}}[\boldsymbol{\theta}] = \text{diag}(\boldsymbol{\theta})$, we have $C_1 = \text{diag}(\mathbf{1})$, where $C_1 \in \mathbb{R}^{N \times N}$. In the case when $C_{\boldsymbol{\theta}}[\boldsymbol{\theta}]$ is a block diagonal matrix with the i -th block $[\theta_1 \ 0 \ \dots \ 0] \in \mathbb{R}^{1 \times n}$, the matrix $C_1 \in \mathbb{R}^{N \times Nn}$ is a block diagonal matrix with the i -th block $[1 \ 0 \ \dots \ 0] \in \mathbb{R}^{1 \times n}$. To compute (32) and (33), we use the recursive expressions (27) and (29).

B. Generalizations

The sensor selection approach presented in Section III is derived on the basis of the output sequence $\{\mathbf{z}_k\}$. In the first sensor selection scenario, explained in Section II, this sequence is computed by simulating the dynamics for the user-selected initial condition. Due to the nonlinearity and non-convexity of the problem, the computed selections of sensor nodes might depend to some extent on the user-selected initial condition. It might happen that the selected sensor locations do not produce satisfactory results when the initial state of the real system significantly differs from the one used to generate the data sequence $\{\mathbf{z}_k\}$. In addition, the solutions of the introduced optimization problems might depend on the initial guesses of the states and sensor locations. If we know the operating range of the initial states of the real system, then we can use this knowledge to properly choose the user-defined initial state for generating the sequence $\{\mathbf{z}_k\}$ and for initializing the optimization problems. In this way, we can reduce the sensitivity of the sensor selection procedure with respect to user-selected initial states. However, in some cases,

due to the lack of knowledge, this is impossible. Our approach for dealing with this problem is to generalize the cost functions in (12) and (13), by including many output and state sequences $\{\mathbf{z}_k\}$ generated for different selections of initial conditions. We can replace the cost functions defined in Section III by the following cost function

$$W[\boldsymbol{\theta}, \mathbf{x}_0] = \|Z - K[\boldsymbol{\theta}]X\|_F^2, \quad (35)$$

$$Z = \begin{bmatrix} \mathbf{z}_{0:L,1} & \mathbf{z}_{0:L,2} & \cdots & \mathbf{z}_{0:L,P} \end{bmatrix}, \quad (36)$$

$$X = \begin{bmatrix} \mathbf{x}_{0:L,1}[\mathbf{x}_{0,1}] & \mathbf{x}_{0:L,2}[\mathbf{x}_{0,2}] & \cdots & \mathbf{x}_{0:L,P}[\mathbf{x}_{0,P}] \end{bmatrix},$$

where $\mathbf{z}_{0:L,i}$ and $\mathbf{x}_{0:L,i}$, $i = 1, 2, \dots, P$, are the output and state sequences, respectively, $K[\boldsymbol{\theta}]$ is defined in (30), and $\|\cdot\|_F^2$ is the Frobenius norm. The sequences $\{\mathbf{z}_{0:L,i}\}$ are obtained by simulating the system dynamics for different initial states. The number P is the total number of user-selected initial states. For example, the initial states can be selected as corners of a convex hull of the expected state-space region from where the state trajectory starts. In some sense, this generalization produces an averaged sensor selection $\hat{\boldsymbol{\theta}}$ that is optimal for a number of initial conditions. The disadvantage of this approach is that the computational complexity of solving the sensor selection problems is increased.

V. NUMERICAL RESULTS

In this section, we present numerical results of applying the developed algorithms to associative memory, Duffing oscillator, and chemical reaction networks. Due to the paper brevity, we mainly focus on networks with stable equilibrium points, and in future work, we will perform tests on networks with limit-cycle dynamics. The simulations are performed on a computer with 16GB RAM and Intel Core i7-8700 processor. The developed codes are available online [40].

A. Sensor Selection for Associative Memory Networks

Associative memory networks can be seen as simplified memory models [48]–[50]. The idea is to select network parameters such that starting from an initial state, representing a perturbed letter or a binary pattern, the network state converges to the equilibrium state, representing the memorized letter or the memorized binary pattern. For brevity, we give a final state-space form of the network, for more background information, see [48]–[50] and references therein. The dynamics of the i -th node has the following form

$$\dot{x}_i = \sum_{j=1}^N \beta_{ij} \sin(x_j - x_i) + \frac{\gamma}{N} \sum_{j=1}^N \sin 2(x_j - x_i), \quad (37)$$

where $x_i \in \mathbb{R}$, $\gamma = 0.8$, $\beta_{ij} \in \mathbb{R}$ is given by $\beta_{ij} = (1/N) \sum_{\omega=1}^p \zeta_i^\omega \zeta_j^\omega$ (Hebb's learning rule), where $\zeta_i^\omega = \pm 1$, $\omega = 1, 2, \dots, p$, and p denotes the number of binary patterns to be memorized. A binary pattern to be memorized is represented by the vector $\zeta^\omega = \text{col}(\zeta_1^\omega, \zeta_2^\omega, \dots, \zeta_N^\omega)$. Let $\zeta^1, \zeta^2, \dots, \zeta^p$ represent p desired binary patterns that need to be memorized by the network. Then for such a selection of binary patterns, we can compute β_{ij} using the previously explained formula, and we can construct the network dynamics (37). We use $N = 25$ and we choose the FE method to discretize the dynamics since as demonstrated in [48] this method produces a small discretization error and the network dynamics is not stiff. We define binary patterns on a 5×5 grid, corresponding to the letters “H”, “T”, and “L” ($p = 3$). The local state trajectories simulated from a perturbed state are shown in Fig. 1(a). In order to be in the transient regime during sensor selection, we use $h = 10^{-3}$ and $L = 21$. The network initial state is the “T” letter that is perturbed by the normal Gaussian noise. Our goal is to estimate this initial state. We test the performance of the method for 40%, 56%, 72%, and 88% of observed nodes (corresponding to $M_{\max} = 10, 14, 18, 22$). For Algorithm 1, we use the less than equal constraints in (12). This is necessary since we have noticed that in the case of equality

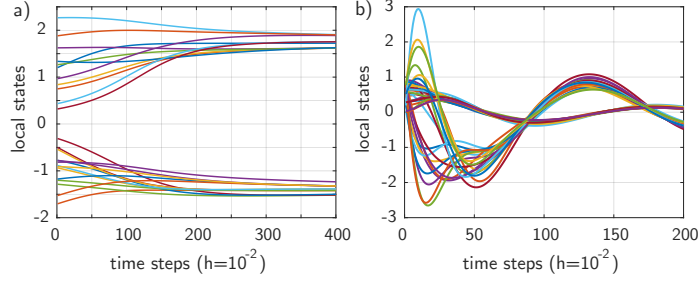


Fig. 1: The state trajectories of a) the associative memory networks and b) Duffing oscillator networks.

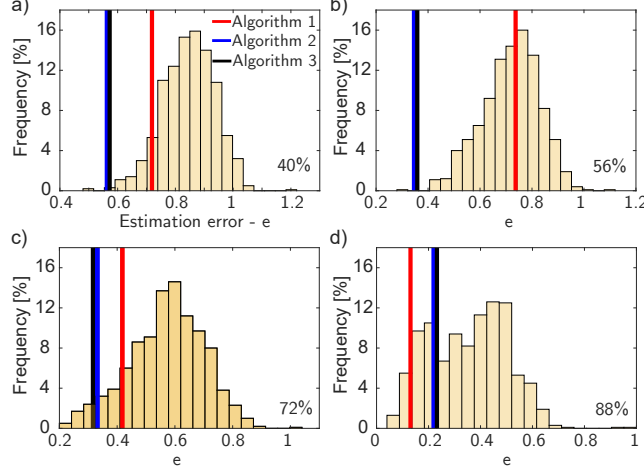


Fig. 2: The estimation errors for associative memory networks. The vertical colored lines correspond to the final estimation errors produced by the developed algorithms and histograms correspond to empirical error distributions obtained by random sensor selection for the fixed fraction of observation nodes. (a) 40%, (b) 56%, 72%, and 88% of observed nodes.

constraints the NOMAD solver occasionally produces infeasible solutions. On the other hand, in the case of Algorithms 2 and 3, we use the equality constraints. For the relaxed problem **P2** in (13), the initial guess of the unknown state is a vector with entries drawn from the Gaussian normal distribution, and the initial guess of the sensor locations is a vector with entries drawn from the uniform distribution on $[0, 1]$. In the relaxed problem, we use the lower and upper bounds equal to -5 and 5 , respectively, and 0 and 1 for the relaxed binary constraints.

To test the performance of the developed algorithms, we compare optimal sensor selections with a sequence of random selections of control nodes. For every random selection, we form the corresponding matrix C , and we estimate the initial state by solving (11). The estimation error is quantified by $e = \|\mathbf{x}_0^{\text{true}} - \hat{\mathbf{x}}_0\|_2 / \|\mathbf{x}_0^{\text{true}}\|_2$, where $\mathbf{x}_0^{\text{true}}$ is the “true” value of the initial state. Figure 2 compares the performance of the developed methods with random sensor selections. We have generated 1000 random selections of sensor nodes. Histograms correspond to empirical distributions of the estimation error produced by random sensor selections. The colored vertical lines denote estimation errors produced by optimal sensor selections computed by the developed algorithms. We can notice that the performance of algorithms varies with a fraction of observed nodes. However, for a fixed fraction of sensor nodes, there is at least one algorithm that produces an error that is smaller than most of the errors produced by the random sensor selection. This is a numerical verification of a relatively good performance of the developed approaches.

B. Sensor Selection for Duffing Oscillator Networks

Duffing oscillators are dynamical systems with a nonlinear spring stiffness $F_s = \eta x - \chi x^3$, where x is the spring displacement, $\eta, \chi \in \mathbb{R}$ are the spring constants, and F_s is the spring force (we assume a softening spring). These oscillators are prototypical models of a large number of dynamical systems, such as electrical circuits, nanomechanical resonators, structural beams, cables, etc. An excellent introduction to the theory and applications of Duffing oscillators is given in [51]. We consider a nonlinear network consisting of damped Duffing oscillators with nonlinear connections:

$$\dot{x}_{i1} = x_{i2}, \quad (38)$$

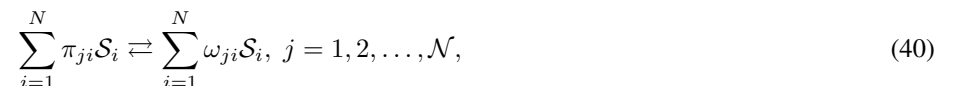
$$\begin{aligned} \dot{x}_{i2} = & -\eta_{ii}x_{i1} + \chi_{ii}x_{i1}^3 - \rho_{ii}x_{i2} - \sum_{j \in \mathcal{N}(i)} \eta_{ij}(x_{i1} - x_{j1}) \\ & + \sum_{j \in \mathcal{N}(i)} \chi_{ij}(x_{i1} - x_{j1})^3 - \sum_{j \in \mathcal{N}(i)} \rho_{ij}(x_{i2} - x_{j2}), \end{aligned} \quad (39)$$

where x_{i1} is the displacement and x_{i2} is the velocity of the i -th oscillator, η_{ij} and χ_{ij} are previously introduced spring constants, ρ_{ij} is a damping parameter, $\mathcal{N}(i)$ is a set of oscillators that are connected to the i -th oscillator. We assume that connections between oscillators are defined by a Geometric Random Graph (GRG). We use the method and codes available in [52] to generate GRGs. GRGs are constructed by randomly placing nodes on a unit square, and connecting nodes according to a user-defined connection radius. In our simulations, we use the radius of $\sqrt{1.44/N}$. The parameters η_{ij} are selected from the uniform distribution defined on the interval $[10, 20]$. On the other hand, parameters χ_{ij}, ρ_{ij} are selected from the uniform distribution defined on $[1, 2]$. The state of the i -th node is $\mathbf{x}^{(i)} = [x_{i1} \ x_{i2}]^T$. Typical local state trajectories simulated from a random initial state and for $N = 20$ are shown in Fig. 1(b).

To test the sensor selection procedure, we construct a Duffing oscillator network with $N = 10$ nodes. Since N is not large, we can compare our methods with an exhaustive search. In this way, we can quantify how far are computed solutions from the most optimal ones. In our simulations, we choose $h = 10^{-3}$ and $L = 101$, such that the collected data belongs to the transient regime. We use the TI method since the FE methods for the Duffing network produces unstable results. The entries of a “true” initial state (to be estimated) are generated from the uniform distribution on the interval $[0, 1]$. An initial guess for this solution is generated using the same principle, however, we make sure that the entries of the guess and “true” initial state are different. The initial guess for the sensor nodes for the relaxed problem **P2** in (13) is generated as a random vector whose entries are selected from the uniform distribution on $[0, 1]$. The principle for selecting other optimization parameters is similar to the principle used to select the optimization parameters for the case of memory networks, for more details, see the codes that are posted online [40]. We test the performance of the methods for 20%, 40%, 60%, and 80% of observed nodes (corresponding to $M_{\max} = 2, 4, 6, 8$). At the same time, we perform the exhaustive search by generating all the possible combinations for fixed fractions of observation nodes, and by estimating an initial state for every combination by solving (11). The results are shown in Fig. 3. We can observe that for every case of the fixed fraction of observed nodes, at least one algorithm produces the estimation error that is relatively close to the most optimal one.

C. Sensor Selection for Chemical Reaction Networks

Due to the length constraints of this manuscript, we provide a brief description of combustion reaction networks, for more details see [9], [53], [54] and references therein. We consider a network consisting of \mathcal{N} chemical reactions



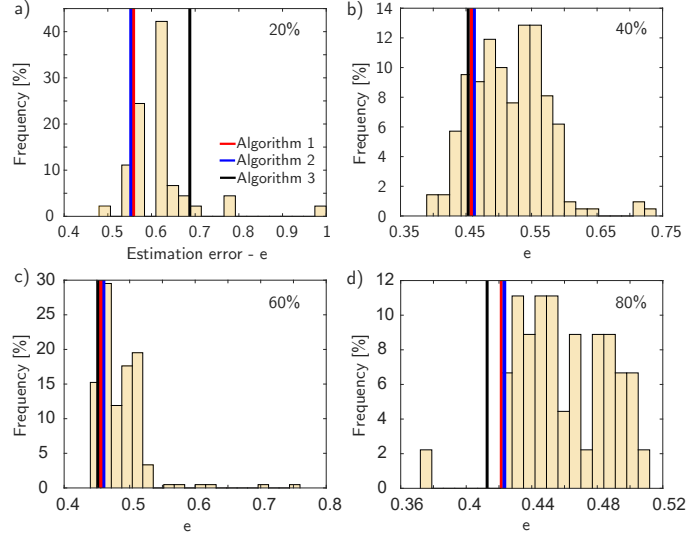


Fig. 3: The estimation errors for the Duffing oscillator networks. The interpretation of vertical lines and histograms is the same as in Fig. 2 and is explained in the caption of Fig. 2. (a) 20%, (b) 40%, 60%, and 80% of observed nodes.

where \mathcal{S}_i , $i = 1, 2, \dots, N$ are chemical species, π_{ji} and ω_{ji} are stoichiometric coefficients. In the state-space description of the chemical reaction networks, the states are chemical species concentrations. The state-space model is

$$\dot{\mathbf{x}}(t) = \Phi \boldsymbol{\lambda}[\mathbf{x}], \quad (41)$$

where $\boldsymbol{\lambda}[\mathbf{x}] = [\lambda_1[\mathbf{x}], \lambda_2[\mathbf{x}], \dots, \lambda_{\mathcal{N}}[\mathbf{x}]]^T$, and $\Phi = [\omega_{ji} - \pi_{ji}] \in \mathbb{R}^{N \times \mathcal{N}}$, $\mathbf{x} = [x_1, x_2, \dots, x_N]$, and λ_j , $j = 1, 2, \dots, \mathcal{N}$ are functions that are polynomials of \mathbf{x} . The local states x_i , $i = 1, 2, \dots, N$ are the concentrations of chemical species. The polynomial functions are defined by

$$\lambda_j[\mathbf{x}] = v_j^f \prod_{i=1}^N x_i^{\pi_{ji}} - v_j^b \prod_{i=1}^N x_i^{\omega_{ji}}, \quad j = 1, 2, \dots, \mathcal{N}, \quad (42)$$

where $v_j^f, v_j^b \in \mathbb{R}_+$ are the forward and backward rate constants computed on the basis of the Arrhenius law. In our simulations, all the network parameters are computed using the Cantera software, for more details, see [9] and references therein.

We consider two combustion chemical reaction networks. The first network is the H_2/O_2 network having 9 species and 27 reactions ($N=9$). The second network is the natural gas combustion network GRI-Mech 3.0 consisting of 53 species and 325 reactions ($N=53$). The dynamics is generated for the temperature of $T = 1473.15[K]$ and the pressure of 1 atmosphere. The local state trajectories of these networks are shown in Fig. 4. To discretize the system dynamics, we used the TI method since the networks are relatively stiff.

Figure 5 and Fig. 6 show the performance of the developed methods for the H_2/O_2 and natural gas combustion networks, respectively. For both networks, the results are generated for $h = 10^{-13}$ and $L = 100$. The principle for selecting other optimization parameters is similar to the principle used to select the optimization parameters for the case of memory networks, for more details, see the codes that are posted online [40]. In the case of the H_2/O_2 network we compare the developed methods with an exhaustive search. In the case of the natural gas combustion network, we compare the methods with 100 random selections of sensor nodes. From Fig. 5 we can observe that the performance of Algorithm 3 is almost ideal for the case of the H_2/O_2 network. On the other hand, from Fig. 6 we can observe that the performance of this algorithm is slightly worse for the case of the natural combustion network. The performance of all three algorithms can be improved by increasing the

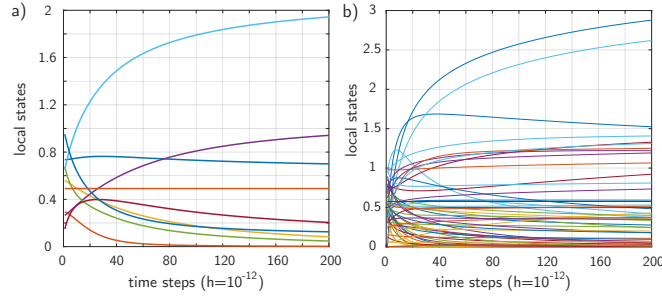


Fig. 4: The local state trajectories of the a) H_2/O_2 and b) natural gas combustion networks.

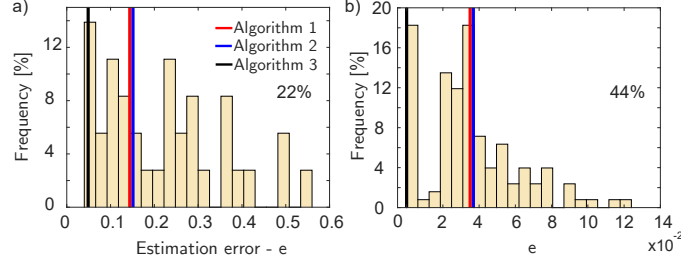


Fig. 5: The estimation errors for the H_2/O_2 networks. The interpretation of vertical lines and histograms is the same as in Fig. 2 and is explained in the caption of Fig. 2. (a) 22% ($M_{\max} = 2$) and (b) 44% ($M_{\max} = 4$) of observed nodes.

length of the observation horizon. However, this performance gain is at the expense of increasing the computational complexity of the developed approaches.

D. Computational Complexity Tests

Finally, we test the computational complexity of the developed methods on the example of the Duffing oscillator network. Due to the many layers of complexity and the fact that we use several nonlinear solvers to compute the solutions, it is challenging to state theoretical bounds on the computational complexity. Consequently, we provide numerical insights into computational complexity. The results are shown in Fig. 7. Here, we used the TI dynamics, $L = 201$, and $h = 10^{-4}$. That is, to compute the solution it is necessary to simulate the dynamics by solving the system of nonlinear equations. The computational complexity times are two orders of magnitude smaller for the FE dynamics. From Fig. 3 we can see that Algorithm 3 produces the lowest computational complexity. This is due to the fact that the **MILP2** problem in (26) only contains $N + 1$ optimization variables. On the other hand, as expected, Algorithm 1 produces the highest computational complexity since it is based on the direct search. The presented computational complexity results can be improved by employing parallel implementations of the used nonlinear solvers.

VI. CONCLUSIONS AND FUTURE WORK

In this manuscript, we have developed methods for sensor selection for nonlinear networks. We performed extensive numerical experiments that demonstrate the good performance of the developed methods. We noticed that the performance of the developed approaches largely depends on the parameters such as the length of the observation horizon, discretization constant, and nonlinear solver parameters. Furthermore, we have noticed that the performance of the developed methods depends on the network types. For example, the best results are obtained for the memory and H_2/O_2 networks. In the case of Duffing oscillator and natural

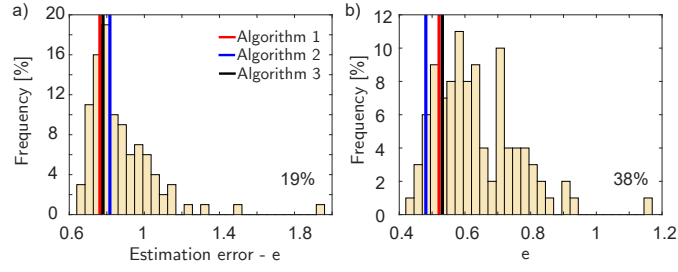


Fig. 6: The estimation errors for the GRI 3.0 networks. The interpretation of vertical lines and histograms is the same as in Fig. 2 and is explained in the caption of Fig. 2. (a) 19% ($M_{\max} = 10$) and (b) 38% ($M_{\max} = 20$) of observed nodes.

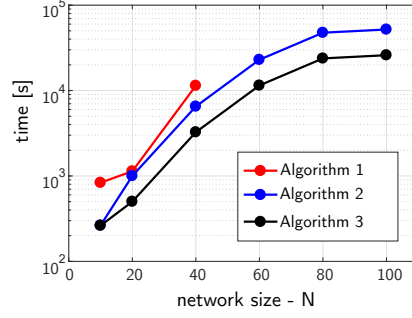


Fig. 7: The computational complexity results of the developed algorithms. The results are obtained for the TI dynamics and the Duffing oscillator networks with $L = 201$ and $h = 10^{-4}$.

gas combustion networks, the performance is slightly worse. The performance can be improved by incorporating higher-order discretization techniques into the developed optimization methods. In future research, we will systematically investigate the influence of network and optimization parameters on the overall performances of the developed methods. On the other hand, due to the nonlinear nature of the problem, optimal sensor selections depend on the output sequence used for optimization. In future research, using the directions explained in Section IV-B, we will generalize the approach by including a number of output and state sequences in the sensor selection problem and we will investigate different options for reducing the computational complexity of the developed methods. In this way, we will further improve the performance of the developed approaches. In addition, we will test our methods on networks exhibiting limit cycle behavior. Finally, the proposed approaches will serve as the basis for the development of closed-loop observers.

REFERENCES

- [1] A. E. Motter, “Networkcontology,” *Chaos*, vol. 25, no. 9, p. 097621, 2015.
- [2] S.-P. Pang, W.-X. Wang, F. Hao, and Y.-C. Lai, “Universal framework for edge controllability of complex networks,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, 2017.
- [3] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, “Controllability of complex networks,” *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.
- [4] Y.-Y. Liu, J.-J. Slotine, and A. Barabási, “Observability of complex systems,” *Proc. Natl. Acad. USA*, vol. 110, no. 7, pp. 2460–2465, 2013.
- [5] M. T. Angulo, C. H. Moog, and Y.-Y. Liu, “A theoretical framework for controlling complex microbial communities,” *Nat. Commun.*, vol. 10, no. 1, pp. 1–12, 2019.
- [6] A. Haber and M. Verhaegen, “Subspace identification of large-scale interconnected systems,” *IEEE Trans. Automat. Contr.*, vol. 59, no. 10, pp. 2754–2759, 2014.
- [7] —, “Sparsity preserving optimal control of discretized pde systems,” *Comput. Methods Appl. Mech. Eng.*, vol. 335, pp. 610–630, 2018.
- [8] —, “Sparse solution of the Lyapunov equation for large-scale interconnected systems,” *Automatica*, vol. 73, pp. 256–268, 2016.

- [9] A. Haber, F. Molnar, and A. E. Motter, "State observation and sensor selection for nonlinear networks," *IEEE Control Netw. Syst.*, vol. 5, no. 2, pp. 694–708, 2017.
- [10] A. Haber and M. Verhaegen, "Moving horizon estimation for large-scale interconnected systems," *IEEE Trans. Automat. Contr.*, vol. 58, no. 11, pp. 2834–2847, 2013.
- [11] —, "Identification of spatially distributed discrete-time state-space models," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 410–415, 2012.
- [12] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, 2014.
- [13] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Control Netw. Syst.*, vol. 3, no. 1, pp. 91–101, 2015.
- [14] C. Letellier, I. Sendiña-Nadal, and L. A. Aguirre, "Nonlinear graph-based theory for dynamical network observability," *Phys. Rev. E*, vol. 98, no. 2, p. 020303, 2018.
- [15] L. A. Aguirre, L. L. Portes, and C. Letellier, "Structural, dynamical and symbolic observability: From dynamical systems to networks," *PLOS ONE*, vol. 13, no. 10, 2018.
- [16] K. Reinschke, *Multivariable Control: A Graph-Theoretic Approach*. Springer, 1988, vol. 41.
- [17] C.-T. Lin, "Structural controllability," *IEEE Trans. Automat. Contr.*, vol. 19, no. 3, pp. 201–208, 1974.
- [18] D. D. Siljak, *Decentralized Control of Complex Systems*. Courier Corporation, 2011.
- [19] M. Doostmohammadian, "Recovering the structural observability of composite networks via Cartesian product," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 133–139, 2020.
- [20] M. T. Angulo, A. Aparicio, and C. H. Moog, "Structural accessibility and structural observability of nonlinear networked systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1656–1666, 2019.
- [21] D. Kahl, P. Wendland, M. Neidhardt, A. Weber, and M. Kschischo, "Structural invertibility and optimal sensor node placement for error and input reconstruction in dynamic systems," *Phys. Rev. X*, vol. 9, no. 4, p. 041046, 2019.
- [22] A. Olshevsky, "On a relaxation of time-varying actuator placement," *IEEE Contr. Syst. Lett.*, vol. 4, no. 3, pp. 656–661, 2020.
- [23] I. S. Klickstein and F. Sorrentino, "Control distance and energy scaling of complex networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 726–736, 2020.
- [24] S. Lall, J. E. Marsden, and S. Glavaški, "A subspace approach to balanced truncation for model reduction of nonlinear control systems," *Int. J. Robust Nonlinear Control.*, vol. 12, no. 6, pp. 519–535, 2002.
- [25] J. Qi, K. Sun, and W. Kang, "Optimal PMU placement for power system dynamic state estimation by using empirical observability Gramian," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2041–2054, 2014.
- [26] M. Serpas, G. Hackebeil, C. Laird, and J. Hahn, "Sensor location for nonlinear dynamic systems via observability analysis and MAX-DET optimization," *Comput. Chem. Eng.*, vol. 48, pp. 105–112, 2013.
- [27] S. A. Nugroho and A. F. Taha, "Sensor placement strategies for some classes of nonlinear dynamic systems via lyapunov theory," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 4551–4556.
- [28] C. Letellier, I. Sendiña-Nadal, E. Bianco-Martinez, and M. S. Baptista, "A symbolic network-based nonlinear theory for dynamical systems observability," *Sci. Rep.*, vol. 8, no. 1, pp. 1–15, 2018.
- [29] I. Sendiña-Nadal and C. Letellier, "Observability of dynamical networks from graphic and symbolic approaches," in *International Workshop on Complex Networks*. Springer, 2019, pp. 3–15.
- [30] Y. Lou and P. D. Christofides, "Optimal actuator/sensor placement for nonlinear control of the Kuramoto-Sivashinsky equation," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 5, pp. 737–745, 2003.
- [31] K. Morris, "Linear-quadratic optimal actuator location," *IEEE Trans. Automat. Contr.*, vol. 56, no. 1, pp. 113–124, 2010.
- [32] M. S. Edalatzaheh and K. A. Morris, "Optimal controller and actuator design for nonlinear parabolic systems," *arXiv:1910.03124*, 2019.
- [33] —, "Optimal actuator design for semilinear systems," *SIAM J. Control Optim.*, vol. 57, no. 4, pp. 2992–3020, 2019.
- [34] 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. Automat. Contr.*, 2019.
- [35] S. Le Digabel, "Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm," *ACM Trans. Math. Softw.*, vol. 37, no. 4, pp. 1–15, 2011.
- [36] S. Sager, "Reformulations and algorithms for the optimization of switching decisions in nonlinear optimal control," *J. Process Control*, vol. 19, no. 8, pp. 1238–1247, 2009.
- [37] S. Sager, M. Jung, and C. Kirches, "Combinatorial integral approximation," *Math. Methods Oper. Res.*, vol. 73, no. 3, p. 363, 2011.
- [38] A. Bürger, C. Zeile, A. Altmann-Dieses, S. Sager, and M. Diehl, "Design, implementation and simulation of an MPC algorithm for switched nonlinear systems under combinatorial constraints," *J. Process Control*, vol. 81, pp. 15–30, 2019.
- [39] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan, "Mixed-integer nonlinear optimization," *Acta Numerica*, vol. 22, pp. 1–131, 2013.

- [40] A. Haber, “Sensor Selection for Nonlinear Networks.” [Online]. Available: <https://git.io/JfMgA>
- [41] A. Alessandri, M. Baglietto, and G. Battistelli, “Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes,” *Automatica*, vol. 44, no. 7, pp. 1753–1765, 2008.
- [42] —, “Receding-horizon estimation for discrete-time linear systems,” *IEEE Trans. Automat. Contr.*, vol. 48, no. 3, pp. 473–478, 2003.
- [43] P. E. Moraal and J. W. Grizzle, “Observer design for nonlinear systems with discrete-time measurements,” *IEEE Trans. Automat. Contr.*, vol. 40, no. 3, pp. 395–404, 1995.
- [44] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, 2009.
- [45] S. Hanba, “On the “uniform” observability of discrete-time nonlinear systems,” *IEEE Trans. Automat. Contr.*, vol. 54, no. 8, pp. 1925–1928, 2009.
- [46] J. Currie, D. I. Wilson, N. Sahinidis, and J. Pinto, “OPTI: Lowering the barrier between open source optimizers and the industrial MATLAB user,” in *Foundations of Computer-Aided Process Operations*, vol. 24, 2012, p. 32.
- [47] L. Grüne and J. Pannek, *Nonlinear Model Predictive Control : Theory and Algorithms*. Springer, 2011.
- [48] A. Haber, S. A. Nugroho, P. Torres, and A. F. Taha, “Control node selection algorithm for nonlinear dynamic networks,” *IEEE Contr. Syst. Lett.*, vol. 5, no. 4, pp. 1195–1200, 2020.
- [49] S. P. Cornelius, W. L. Kath, and A. E. Motter, “Realistic control of network dynamics,” *Nat. Commun.*, vol. 4, no. 1, pp. 1–9, 2013.
- [50] T. Nishikawa, Y.-C. Lai, and F. C. Hoppensteadt, “Capacity of oscillatory associative-memory networks with error-free retrieval,” *Phys. Rev. Lett.*, vol. 92, no. 10, p. 108101, 2004.
- [51] I. Kovacic and M. J. Brennan, *The Duffing Equation: Nonlinear Oscillators and Their Behaviour*. John Wiley & Sons, 2011.
- [52] A. Taylor and D. J. Higham, “CONTEST: A controllable test matrix toolbox for MATLAB,” *ACM Trans. Math. Softw.*, vol. 35, no. 4, pp. 1–17, 2009.
- [53] S. Turns, *An Introduction to Combustion: Concepts and Applications*. McGraw-Hill, 1996.
- [54] N. N. Smirnov and V. F. Nikitin, “Modeling and simulation of hydrogen combustion in engines,” *Int. J. Hydrog. Energy*, vol. 39, no. 2, pp. 1122 – 1136, 2014.