

# Optimal reporter placement in sparsely measured genetic networks using the Koopman operator

Aqib Hasnain, Nibodh Boddupalli, and Enoch Yeung

**Abstract**—Optimal sensor placement is an important yet unsolved problem in control theory. In biological organisms, genetic activity is often highly nonlinear, making it difficult to design libraries of promoters to act as reporters of the cell state. We make use of the Koopman observability gramian to develop an algorithm for optimal sensor (or reporter) placement for discrete time nonlinear dynamical systems to ease the difficulty of design of the promoter library. This ease is enabled due to the fact that the Koopman operator represents the evolution of a nonlinear system linearly by lifting the states to an infinite-dimensional space of observables. The Koopman framework ideally demands high temporal resolution, but data in biology are often sampled sparsely in time. Therefore we compute what we call the temporally *fine-grained* Koopman operator from the temporally *coarse-grained* Koopman operator, the latter of which is identified from the sparse data. The optimal placement of sensors then corresponds to maximizing the observability of the fine-grained system. We demonstrate the algorithm on a simulation example of a circadian oscillator.

## I. INTRODUCTION

Spectral methods have been increasingly popular in data-driven analysis of nonlinear dynamical systems. Recently, researchers working in Koopman operator theory have shown that it is possible to identify and learn the fundamental modes for a nonlinear dynamical system from data [16], [17]. This operator, originally defined nearly 100 years ago by Koopman [8], is a linear infinite dimensional operator that fully describes the underlying nonlinear dynamical system. Identifying Koopman operators from data has become computationally tractable, largely due to advances integrating machine learning and deep learning to generate novel, efficient representations of observable subspaces for the Koopman operator [13], [28].

In many high-dimensional nonlinear systems, typically it is not physically or economically feasible to measure every state with the resolution specified by a fine-grained temporal model. For example, the bacteria *E. coli* have approximately 4400 genes, making both spatially and temporally fine data collection nearly impossible. On one hand, high-coverage omics measurements provide a system-level view of all gene activity, but prohibitive costs and the laborious and destructive nature of sampling make it difficult to resolve dynamics at a high temporal resolution. On the other hand, fluorescently tagged genes can be measured at the second

to minutes timescale, to profile bursty RNA dynamics and protein expression. Knowing which genes to tag with fluorescent markers is critical, since not every gene can be simultaneously tagged. This challenge motivates the need for algorithmic data-driven approaches which allow the user (e.g. biologists) to know *a priori* which genes should be sampled. Finally, is it possible to design a nonlinear observer that rather than measuring a single gene or a single node in the network, fuses the state of a select set of biomarker genes to report out an aggregate cellular state of the system? The fundamental question is how to use metrics for nonlinear observability to design observers or optimize sensor placement. Sinha et al. presented a systematic framework based on linear transfer operators for the optimal placement of sensors and actuators for control of nonequilibrium dynamics [20].

Koopman operators have been used to characterize observability of a nonlinear system [23], [24]. Yeung et al. formulated the Koopman gramian and showed they can be used to quantify controllability and observability and lend insight for the underlying nonlinear dynamical system [29]. This recent development of the Koopman gramians can advance the important and unsolved problem of optimal sensor placement in control theory. The Koopman framework embeds nonlinear dynamics in a linear framework for optimal nonlinear estimation and control [1], [2], [9]. For sensor placement search spaces that are reasonable in size, there are model-based solutions using optimal experiment design [3], [7], information theoretic and Bayesian criteria [4], [10], [11], [15], [18]. There is a need to develop purely data-driven methods for determining optimal sensor placement. Manohar et al. explored optimized sparse sensor placement for signal reconstruction based on a tailored library of features extracted from training data [14]. In [19], Sharma et al. extended the transfer operator based approach for optimal sensor placement, providing a probabilistic metric to gauge coverage under uncertain conditions. Fontanini et al. presented a data driven sensor placement algorithm based on a dynamical systems approach, utilizing the Perron-Frobenius operator [5]. Our framework provides a method to determine optimal sensor placement, even in the presence of noisy and temporally sparse data using Koopman operator theory.

In this paper, we develop an algorithm for optimizing sensor placement from sparsely sampled time-series data. We use the Koopman observability gramian, developed by Yeung et al. [29], to maximize the observability of the underlying discrete time nonlinear dynamical system. Section II introduces the Koopman operator formulation and Section

A. Hasnain and N. Boddupalli are with the Department of Mechanical Engineering, University of California, Santa Barbara aqib@ucsb.edu, nibodh@ucsb.edu

E. Yeung is with the Department of Mechanical Engineering, Center for Control, Dynamical Systems, and Computation, and Biomolecular Science and Engineering, University of California, Santa Barbara eyeung@ucsb.edu

III introduces the notion of a Koopman observability gramian [29]. In Section IV, we show how to compute the temporally fine-grained Koopman operator from the temporally coarse-grained Koopman operator, which is learned from data that are temporally sparse. In the case of noisy data, a closed form expression for the error in computing the temporally coarse-grained Koopman operator is derived. In Section IV-B, we present a novel algorithm for optimal sensor design and placement. Finally, the algorithm is illustrated with a simulation example.

## II. KOOPMAN OPERATOR FORMULATION

We briefly introduce Koopman operator theory, as we will use it extensively for the sensor placement problem. Consider a discrete time open-loop nonlinear system of the form

$$\begin{aligned} x_{t+1} &= f(x_t) \\ y_t &= h(x_t) \end{aligned} \quad (1)$$

with  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is analytic and  $h \in \mathbb{R}^p$ . The Koopman operator of (1),  $\mathcal{K} : \mathcal{F} \rightarrow \mathcal{F}$ , is a linear operator that acts on observable functions  $\psi(x_k)$  and propagates them forward in time as

$$\psi(x_{t+1}) = \mathcal{K}\psi(x_t). \quad (2)$$

Here  $\mathcal{F}$  is the space of observable functions that is invariant under the action of  $\mathcal{K}$ .

*Assumption 1:* Given system (1), we suppose that  $y_k = h(x_k) \in \mathcal{F}$  and that  $h \in \text{span}\{\psi_1, \psi_2, \dots\}$ . Then the output  $y_t$  can be expressed as

$$y_t = h(x_t) = W_h \psi(x_t) \quad (3)$$

where the output matrix  $W_h \in \mathbb{R}^{p \times n_L}$ ,  $n_L \leq \infty$ . We make this strong assumption since the structure of  $W_h$  will be manipulated to achieve optimal sensor placement.

Throughout the paper, we take observable functions which are state-inclusive, i.e.

$$\psi(x) = (x, \varphi(x)) \quad (4)$$

where  $\varphi \in \mathbb{R}^{n_L - n}$  are continuous functions in  $\mathcal{F}$ .

## III. KOOPMAN OBSERVABILITY GRAMIAN

The observability matrix of the transformed system may be obtained by showing how the Koopman operator maps initial conditions  $x_0$  to  $y$  [29]. Using equations (2) and (3), we have

$$y_t = W_h \mathcal{K}^t \psi(x_0) \quad (5)$$

Therefore,  $W_h \mathcal{K}^t : \mathbb{R}^{n_L} \rightarrow \mathbb{R}^p$  is the transformation that maps  $\psi(x_0)$  to  $y_t$ . Given an initial condition  $\psi(x_0) \in \mathbb{R}^{n_L}$ , the energy of the output  $y_t$  is given by

$$\begin{aligned} \|y\|^2 &= \sum_n \langle y_t, y_t \rangle \\ &= \sum_n \psi(x_0)^\top (\mathcal{K}^t)^\top W_h^\top W_h \mathcal{K}^t \psi(x_0) \\ &= \sum_n \psi(x_0)^\top X_o^\psi \psi(x_0) \end{aligned} \quad (6)$$

where  $\langle \cdot, \cdot \rangle$  represents the inner product and as can be seen in the last equality of (6), the Koopman observability gramian is defined as

$$X_o^\psi = \sum_{t=0}^{\infty} (\mathcal{K}^t)^\top W_h^\top W_h \mathcal{K}^t \quad (7)$$

and is an  $n_L \times n_L$  matrix. The observability gramian can be obtained as a solution of following matrix Lyapunov equation

$$\mathcal{K}^\top X_o^\psi \mathcal{K} - X_o^\psi = -W_h^\top W_h.$$

The Koopman observability gramian quantifies the observability of the function  $\psi(x)$ . More importantly, when  $\psi(x)$  includes observable functions related to the local observability of the underlying nonlinear system (1), the Koopman observability gramian retains that information [29].

## IV. SENSOR PLACEMENT FROM TEMPORALLY SPARSE DATA

### A. Fine-grained models from coarse-grained models

We consider the scenario where high-resolution measurements of all genes in a single cell are infrequently sampled. This is a common scenario when tracking the state of biological, cyber-physical, and social networks. Exhaustive measurement of every state in the system is expensive (and often manual) and thus can only be performed infrequently.

Consider the case where the system in (1) is a biomolecular reaction network evolving with unknown governing equations. The precise functional form and parameters of  $f$  are typically considered unknown. In some settings, *a priori* knowledge of the biomolecular reaction network can be utilized to bootstrap the modeling problem [27], [30]. We consider a data-driven operator theoretic approach, using the method of Koopman briefly introduced in Section II.

The discrete time Koopman representation for the system (1) is

$$\psi(x_{t+1}) = K \psi(x_t) \quad (8)$$

where the matrix  $K \in \mathbb{R}^{n_L \times n_L}$  is a finite dimensional approximation of the exact Koopman operator  $\mathcal{K}$  and  $\psi(x_k) \in \mathbb{R}^{n_L}$ . We suppose that full-state measurements are made available for  $x_t, x_{t+N}$ , with enough biological replicates that the *temporally coarse-grained* (approximate) Koopman operator is identifiable via the optimization problem

$$\min_{K_N} \|\Psi(X_f) - K_N \Psi(X_p)\|$$

where

$$\begin{aligned} \Psi(X_f) &\equiv [\psi(x_{t+N}(\omega_1)) \quad \dots \quad \psi(x_{t+N}(\omega_R))], \\ \Psi(X_p) &\equiv [\psi(x_t(\omega_1)) \quad \dots \quad \psi(x_t(\omega_R))]. \end{aligned}$$

and  $\omega_R$  represents the number of replicates. In the presence of sparse and noisy data, [21] showed that the Koopman learning problem can be formulated as a robust optimization problem, which is equivalent to a specific regularized learning problem in which the LASSO penalty parameter corresponds to the upper bound on the noise i.e. the maximum Frobenius norm of the noise. We will suppose, for simplicity

of exposition of the technique, that an exact Koopman operator for the coarse-time step mapping  $t$  to  $t+N$  is either known or obtained directly from data satisfying

$$\psi(x_{t+N}) = K_N \psi(x_t). \quad (9)$$

Because of linearity of the Koopman operator, we know that the *temporally fine-grained* Koopman operator  $K$  satisfies

$$\psi(x_{t+1}) = K \psi(x_t)$$

and most importantly,

$$K = K_N^{1/N}. \quad (10)$$

When the Koopman observable function includes the state as an element, this relationship allows the recovery of the fine-grained governing equations for  $f$  directly from a temporally coarse-grained Koopman operator (and the corresponding data). To see this, take the state-inclusive observable functions (4) and partition the Koopman equation accordingly as

$$\begin{bmatrix} x_{t+1} \\ \varphi(x_{t+1}) \end{bmatrix} = \begin{bmatrix} K_{xx} & K_{x\varphi} \\ K_{\varphi x} & K_{\varphi\varphi} \end{bmatrix} \begin{bmatrix} x_t \\ \varphi(x_t) \end{bmatrix}. \quad (11)$$

Since the Koopman operator satisfies

$$K \psi(x_t) = \psi(f(x_t)) \quad (12)$$

for each row, then in particular, the upper half of the Koopman equation satisfies

$$x_{t+1} = K_{xx}x_t + K_{x\varphi}\varphi(x_t) = f(x_t). \quad (13)$$

This provides a powerful scheme for estimating the governing equations of a fine-grained time-evolving biological process from sparse or coarse-grained temporal measurements, so long as the coarse-grained time measurement is a product of regularly spaced intervals of time in the fine-grained representation. Again, since RNAseq and proteomic measurements often provide full-state measurements of a network, this in theory can provide sufficient information to recover the Koopman operator, even in the presence of noise [21]. The key insight and property leveraged is the linearity of the lifted Koopman representation. One would not be able to obtain the fine-grained dynamics of the governing equations from a coarse grained representation of the governing equations as it is generally not feasible to compute the  $n^{\text{th}}$  root of a  $n$ -layered function composition. Specifically, note that the  $N$ -step map for the underlying governing dynamics of system (9) is given as

$$x_{t+N} = f^{(n)}(x_t) = f \circ f \circ \dots \circ f(x_t) \equiv f_N(x_t). \quad (14)$$

Given an arbitrary nonlinear function  $f_N(x_t)$  that is the  $N^{\text{th}}$  composition of  $f(x_t)$ , there is no general way to obtain the underlying  $f(x_t)$ . However, by using the Koopman operator lifting framework, we can express the governing equations with linear coordinates, which allows us to consider computing the  $N^{\text{th}}$  root to obtain the single-step map from the  $N$ -step map.

Although, in general, the matrix root always exists, we note that we may not always obtain the desired fine-grained

$K$  from  $K_N$  due to there being multiple solutions to matrix roots. Yue et al [31] showed that similarly the matrix logarithm raises a concept of system aliasing. They describe the scenario where there might be multiple fine-grained systems which give the same coarse-grained system. In the case that multiple fine-grained Koopman operators exist, our method can be applied to each operator. We can distinguish which operator is the "correct" operator by collecting a few data points at a fine-grained temporal resolution and evaluating the predictive accuracy of the fine-grained Koopman operator models.

In the presence of noise, we approximate the fine-grained discrete time Koopman operator  $K$  from the coarse-grained discrete time Koopman operator  $K_N$  as

$$\begin{aligned} \hat{K} &= \hat{K}_N^{(1/N)} = (K_N + \epsilon(x))^{1/N} \\ &= \sum_{k=0}^{\infty} \binom{1/N}{k} K_N^{(1/N-k)} \epsilon(x)^k \\ &= \hat{K}_N^{1/N} + \frac{1}{N} K_N^{(1/N-1)} \epsilon(x) \\ &\quad + \frac{\frac{1}{N}(\frac{1}{N}-1)}{2!} K_N^{(1/N-2)} \epsilon(x)^2 + \dots \end{aligned} \quad (15)$$

where the last equality follows from Newton's generalization of the binomial theorem [12]. Here we assume that  $\epsilon(x)$  is bounded as in [6] for all  $x \in \mathcal{M} \subseteq \mathbb{R}^n$ . A closed form expression of the error term  $\epsilon(x)$  is found by noting that

$$\epsilon(x) = \hat{K}_N - K_N.$$

Then we have

$$\epsilon(x) \Psi(X_p) = (\hat{K}_N - K_N) \Psi(X_p) = \hat{\Psi}(X_f) - \Psi(X_f)$$

$$\epsilon(x) \Psi(X_p) \Psi(X_p)^\dagger = (\hat{\Psi}(X_f) - \Psi(X_f)) \Psi(X_p)^\dagger.$$

giving the closed form expression of the error as

$$\epsilon(x) = (\hat{\Psi}(X_f) - \Psi(X_f)) \Psi(X_p)^\dagger. \quad (16)$$

Once we obtain the one-step Koopman operator, notice that the Koopman invariant subspace of observable functions is the same as the  $N$ -step operator. We suppose, mirroring the scenario presented with transcriptomic and proteomic measurements, that the state is measured completely, in this setting. The precise coverage of the entire transcriptome and proteome is often a subject of debate, but relative to the spatial sparsity of fluorescence based readout approaches, we shall assume for our purposes that the full state of the network is measured sparsely.

The state-output equations of the coarse-grained system can then be written as

$$\begin{aligned} x_{t+N} &= f(x_t) \\ y_t &= x_t \end{aligned} \quad (17)$$

and thus the corresponding Koopman equation can be written as

$$\begin{aligned} \psi(x_{t+N}) &= K_N \psi(x_t) \\ y_t &= P_x \psi(x_t) \end{aligned} \quad (18)$$

where

$$P_x = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$$

is the projection matrix that extracts the state observable from the vector observable  $\psi(x_t)$ .

### B. Fine-Grained Sensor Placement via Optimal Observability

Often times, it is not physically or economically feasible to measure every state with the resolution specified by a fine-grained temporal model. We seek to develop an algorithm for identifying the design and placement of reporters that maximizes the observability of the underlying nonlinear system, as well as the corresponding Koopman representation. For this task, we find it convenient to pose this problem using the Koopman gramian as defined in Section III. Specifically, we seek to construct an output observer for the fine-grained dynamical system (1) given full-state sparse temporal measurements at  $t, t+N, t+jN$  in sufficient frequency to recover the temporally coarse-grained Koopman operator  $K_N$ , so that it is possible to compute the fine-grained Koopman operator  $K = K_N^{(1/N)}$ . We suppose that the corresponding Koopman representation with output equation is thus written as

$$\begin{aligned} \psi(x_{t+1}) &= K\psi(x_t) \\ y_t &= W_h\psi(x_t). \end{aligned} \quad (19)$$

We seek to maximize the output energy  $\|y_t\|^2$  for an initial condition  $x_0$  at a time instant  $t$  i.e. solve the nonlinear optimization problem

$$\max_{h(x) \in \mathcal{L}^2(\mathcal{M})} \|y(t_j)\|^2 \quad (20)$$

for all initial conditions  $x_0$  with  $\|x_0\| \leq 1$ . This is an optimization problem of a nonlinear function space (i.e. an uncountably infinite dimensional space) and is generally intractable. However, if we were to find a basis for  $h(x)$ , we could express the problem in terms of a linear combination of the basis functions, which would yield a convex formulation of the problem. This is precisely what we can do using the spectral properties of the Koopman operator representation. Following the formulation given in Section III, the system in (19) has Koopman observability gramian

$$X_{o,f}^\psi = \sum_{j=0}^{t_N} (K^j)^\top W_h^\top W_h (K^j) \quad (21)$$

where the subscript  $f$  is used to distinguish the fine-grained system from coarse-grained.

We want to identify the optimal sensor placement that informs the design of optimal observers. Utilizing the Koopman observability gramian,  $X_{o,f}$ , as defined in (21), the output energy of system (19) is written as

$$\|y_{t_N}\|^2 = \sum_{j=0}^{t_N} \psi(x_0)^\top (K^j)^\top W_h^\top W_h (K^j) \psi(x_0). \quad (22)$$

Our goal is to now maximize the output energy (22) of the lifted system up at time  $t$  with the output matrix  $W_h$  as the

decision variable. If the output energy of the lifted system is maximized, then by proxy the output energy of the original nonlinear system is maximized.

For the purposes of this paper, we will suppose that we construct an observable function basis that results in a diagonalizable Koopman operator. The subsequent presentation can be generalized for Koopman operators that only admit a Jordan decomposition, but for simplicity of exposition, we consider the case of the diagonalizable Koopman operator.

*Assumption 2:* We suppose that  $\psi(x)$  and  $K$  are provided or trained during the learning process to admit a diagonalizable  $K$ .

Thus, an eigendecomposition of  $K$  gives

$$KV = V\Lambda$$

where  $V$  is an  $n_L \times n_L$  matrix of eigenvectors. The  $n_L \times n_L$  matrix  $\Lambda$  is a diagonal matrix whose components are the eigenvalues  $\lambda$  of the Koopman operator,  $K$ . The eigenfunctions of  $K$  are then written as

$$\phi(x_0) = V^{-1}\psi(x_0).$$

where  $\phi \in \mathbb{R}^{n_L}$ . Since (22) has a symmetric form, let us deal with the right half of this equation. We have that

$$\begin{aligned} W_h K^j \psi(x_0) &= W_h V \Lambda^j V^{-1} \psi(x_0) \\ &= W_h V \Lambda^j V^{-1} V \phi(x_0) \\ &= W_h V \Lambda^j \phi(x_0). \end{aligned}$$

The output energy can now be written in terms of the Koopman eigenfunctions as

$$\|y_{t_N}\|^2 = \sum_{j=0}^{t_N} \left[ \phi(x_0)^\top \Lambda^j V^\top W_h^\top W_h V \Lambda^j \phi(x_0) \right] \quad (23)$$

The optimization problem (20) can now be formulated as

$$\mathcal{J} = \max_{W_h} \sum_{j=0}^{t_N} \left[ \phi(x_0)^\top \Lambda^j V^\top W_h^\top W_h V \Lambda^j \phi(x_0) \right] \quad (24)$$

with  $\|W_h^\top W_h\|_2 \leq C$ . The upper bound  $C$  would vary between biological experiments and should be identified directly from data.

By picking out the  $p$  ( $\leq n_L$ ) most observable modes of the system such that we can ensure the collection of measurements which correspond to maximal energy. If we define  $W_h$  as

$$W_h \triangleq [I_{p \times p} \quad 0] V^{-1}$$

the argument of (24) becomes

$$\begin{aligned} \sum_{j=0}^{t_N} \left( \phi(x_0)^\top \Lambda^j V^\top (V^{-1})^\top \begin{bmatrix} I_{p \times p} \\ 0 \end{bmatrix} \begin{bmatrix} I_{p \times p} & 0 \end{bmatrix} V^{-1} V \Lambda^j \phi(x_0) \right) \\ = \sum_{j=0}^{t_N} \left( \phi(x_0)^\top \text{diag}(\lambda_1^{2j}, \lambda_2^{2j}, \dots, \lambda_p^{2j}, 0, \dots, 0) \phi(x_0) \right) \end{aligned}$$

where  $\lambda_1$  through  $\lambda_p$  are the  $p$  maximum eigenvalues of  $K$ . The maximum output energy comes from a choice of  $W_h$  that depends on the eigenvectors of the Koopman operator.

1) *Example (Circadian oscillator)*: To illustrate our sensor placement algorithm, we consider a model of a circadian oscillator, see Vilar et al. [25], that involves an activator  $A$  and a repressor  $R$ . Both  $A$  and  $R$  are transcribed into  $mRNA$  and subsequently translated into protein. Since  $A$  can bind to both  $A$  and  $R$  promoters, it increases their transcription rates.  $R$  acts as a negative element by hindering  $A$ . The deterministic dynamics are given by the following reaction rate equations

$$\begin{aligned}
\dot{D}_A &= \theta_A D'_A - \gamma_A D_A A \\
\dot{D}_R &= \theta_R D'_R - \gamma_R D_R A \\
\dot{D}'_A &= \gamma_A D_A A - \theta_A D'_A \\
\dot{D}'_R &= \gamma_R D_R A - \theta_R D'_R \\
\dot{M}_A &= \alpha'_A D'_A + \alpha_A D_A - \delta_{MA} M_A \\
\dot{A} &= \beta_A M_A + \theta_A D'_A + \theta_R D'_R \\
&\quad - A(\gamma_A D_A A + \gamma_R D_R + \gamma_C R + \delta_R) \\
\dot{M}_R &= \alpha'_R D'_R + \alpha_R D_R - \delta_{MR} M_R \\
\dot{R} &= \beta_R M_R - \gamma_C A R + \delta_A C - \delta_R R \\
\dot{C} &= \gamma_C A R - \delta_A C.
\end{aligned} \tag{25}$$

Extended dynamic mode decomposition (EDMD) [26] is used to compute the finite-dimensional approximation of the Koopman operator,  $K_N$ , for a coarse time step. A dictionary of state-inclusive observable functions,  $\Psi$ , is constructed using up to second-order polynomials. Often in biological systems, Hill function type nonlinearities appear in the dynamics. Even in these cases, the dictionary of polynomial functions should capture the dynamics well, according to the Weierstrass Approximation theorem, which states that any continuous function on a closed and bounded interval can be uniformly approximated on that interval by polynomials. [22]. Therefore, as long as the non-polynomial linearity is continuous, we expect that this dictionary of polynomials will result in accurate predictions, although the representation may not be as low dimensional as a representation drawn from a more efficient encoding [13], [28].

In this example, initial conditions were chosen such that the trajectories converge to a limit cycle. From the coarse-grained Koopman operator obtained from simulation data, the fine-grained Koopman operator,  $K$ , is computed using the scheme outlined in Section IV. Solving the optimization problem (24), we can identify the optimal sensor placement. Choosing  $p$ , the number of rows in  $W_h$ , to be  $p = 20$ , we get the output matrix structure as seen in figure 1. A total of 55 observable functions were used which correspondingly sets the number of columns in the output matrix,  $W_h$ . The output matrix has a sparse structure with most elements of the matrix nearly zero. Using the criteria that the 1-norm of the columns of  $W_h$  determine the most *active* states of the observable coordinates, we can determine optimal sensor placement. Using this criteria, the most active dynamics are  $M_A C$ ,  $M_R C$ ,  $A C$ ,  $R^2$ ,  $R C$ , and  $C^2$  for a single initial condition where the trajectories converge to limit cycles. Figure

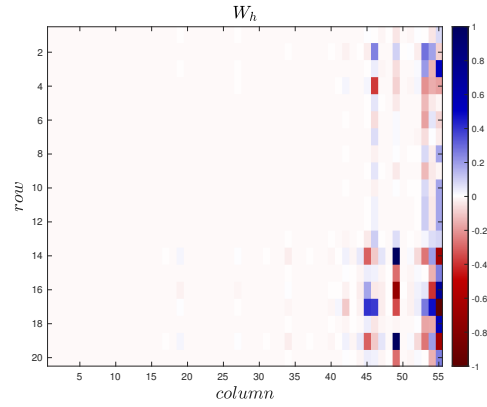


Fig. 1. Sparse structure of the output matrix  $W_h$  with  $p = 20$  for the circadian oscillator simulation.

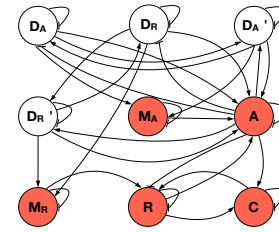


Fig. 2. Network architecture of the circadian oscillator model in (25). Arrows indicate activation, while bars indicate repression or degradation. Highlighted in red are the states which have the most active dynamics in the observable coordinates. Note that these active states were taken from the single initial condition used to produce figure 1.

2 shows the entire network architecture of the circadian oscillator. The states highlighted in red are the active states and correspondingly are where the algorithm would dictate sensors should be placed. Figure 3 shows how frequently a state appears as an active state in the observable coordinates over 20 different initial conditions.

From this analysis, a nonlinear observable can be designed. For example, the state  $C^2$  is highly active in the observable basis, therefore a nonlinear observer can be designed where a  $C$  molecule binds with another  $C$  molecule and integrated to obtain the output. We can then use this observer (and other observers) to act as a reporter for the cell state. This can enable rapid experimentation in synthetic biology since there would no longer be a need to collect expensive full state proteomics and transcriptomics data at a low temporal resolution. We can collect partial state measurements from states of interest at a high temporal resolution. The Koopman method thus can identify critical genes that serve as cell state biomarkers. These biomarkers provide a link between internal dynamics and observed phenotypes.

## V. CONCLUSION

In this work, we developed an algorithm for optimal sensor placement from sparsely sampled time-series data of discrete time nonlinear systems. The optimal sensor

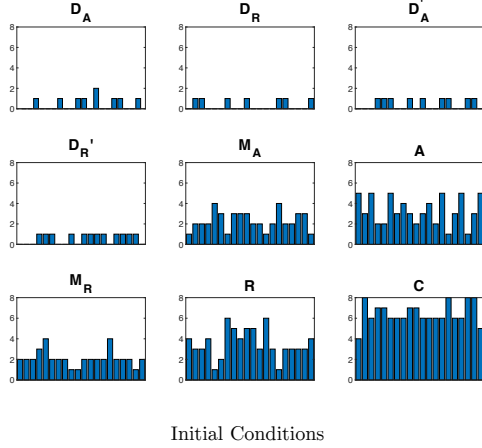


Fig. 3. Histogram showing the frequency of a state of (25) being in the 10 most active states of the observables over 20 different initial conditions.

placement algorithm was formulated as maximizing the observability of a dynamical system or genetic network in the Koopman lifted space. We compute the temporally *fine-grained* Koopman operator from the temporally *coarse-grained* Koopman operator, the latter of which is identified directly from sparse biological data. In the case of noisy data, a closed form expression for the error in the coarse-grained Koopman operator is derived. Finally, we have illustrated the optimal sensor placement method on a simulation example of a circadian oscillator. This method can be utilized in the context of developing bacterial sensors where the design of a library of promoters is now informed by the sensor placement algorithm.

#### ACKNOWLEDGEMENTS

The authors would also like to thank Professor Igor Mezic for insightful discussions. This material is based on work supported by DARPA and AFRL under contract number DEAC0576RL01830. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Defense Advanced Research Project Agency, the Department of Defense, or the United States government.

#### REFERENCES

- [1] I. Abraham, G. De La Torre, and T. D. Murphey. Model-based control using koopman operators. *arXiv preprint arXiv:1709.01568*, 2017.
- [2] H. Arbabi, M. Korda, and I. Mezic. A data-driven koopman model predictive control framework for nonlinear flows. *arXiv preprint arXiv:1804.05291*, 2018.
- [3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [4] W. F. Caselton and J. V. Zidek. Optimal monitoring network designs. *Statistics & Probability Letters*, 2(4):223–227, 1984.
- [5] A. D. Fontanini, U. Vaidya, and B. Ganapathysubramanian. A methodology for optimal placement of sensors in enclosed environments: A dynamical systems approach. *Building and Environment*, 100:145–161, 2016.
- [6] C. A. Johnson and E. Yeung. A class of logistic functions for approximating state-inclusive koopman operators. *2018 Annual American Control Conference (ACC)*, 2018.

- [7] S. Joshi and S. Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2009.
- [8] B. O. Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315318, 1931.
- [9] M. Korda and I. Mezic. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149160, 2018.
- [10] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(Feb):235–284, 2008.
- [11] D. V. Lindley et al. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.
- [12] C.-s. Liu. The essence of the generalized newton binomial theorem. *Communications in Nonlinear Science and Numerical Simulation*, 15(10):2766–2768, 2010.
- [13] B. Lusch, J. N. Kutz, and S. L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):4950, 2018.
- [14] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton. Data-driven sparse sensor placement for reconstruction. *arXiv preprint arXiv:1701.07569*, 2017.
- [15] L. Paninski. Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17(7):1480–1507, 2005.
- [16] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142161, 2016.
- [17] C. W. Rowley, I. Mezic, S. Bagheri, P. Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115, 2009.
- [18] P. Sebastiani and H. P. Wynn. Maximum entropy sampling and optimal bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):145–157, 2000.
- [19] H. Sharma, U. Vaidya, and B. Ganapathysubramanian. A transfer operator methodology for optimal sensor placement accounting for uncertainty. *arXiv preprint arXiv:1812.10541*, 2018.
- [20] S. Sinha, U. Vaidya, and R. Rajaram. Operator theoretic framework for optimal placement of sensors and actuators for control of nonequilibrium dynamics. *Journal of Mathematical Analysis and Applications*, 440(2):750–772, 2016.
- [21] S. Sinha and E. Yeung. On computation of koopman operator from sparse data. *arXiv:1901.03024*, 2019.
- [22] M. H. Stone. The generalized weierstrass approximation theorem. *Mathematics Magazine*, 21(4):167, 1948.
- [23] A. Surana and A. Banaszuk. Linear observer synthesis for nonlinear systems using koopman operator framework. *IFAC-PapersOnLine*, 49(18):716723, 2016.
- [24] U. Vaidya. Observability gramian for nonlinear systems. *2007 46th IEEE Conference on Decision and Control*, 2007.
- [25] J. M. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.
- [26] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A datadriven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):13071346, 2015.
- [27] E. Yeung, A. J. Dy, K. B. Martin, A. H. Ng, D. Del Vecchio, J. L. Beck, J. J. Collins, and R. M. Murray. Biophysical constraints arising from compositional context in synthetic gene networks. *Cell systems*, 5(1):11–24, 2017.
- [28] E. Yeung, S. Kundu, and N. Hodas. Learning deep neural network representations for koopman operators of nonlinear dynamical systems. *arXiv preprint arXiv:1708.06850*, 2017.
- [29] E. Yeung, Z. Liu, and N. O. Hodas. A koopman operator approach for computing and balancing gramians for discrete time nonlinear systems. *2018 Annual American Control Conference (ACC)*, 2018.
- [30] E. Yeung, A. Ng, J. Kim, Z. Z. Sun, and R. M. Murray. Modeling the effects of compositional context on promoter activity in an e. coli extract based transcription-translation system. In *53rd IEEE Conference on Decision and Control*, pages 5405–5412. IEEE, 2014.
- [31] Z. Yue, J. Thunberg, L. Ljung, Y. Yuan, and J. Goncalves. Systems aliasing in dynamic network reconstruction: Issues on low sampling frequencies. *arXiv preprint arXiv:1605.08590*, 2016.