

PAPER

Estimate the spectrum of affine dynamical systems from partial observations of a single trajectory data

To cite this article: Jiahui Cheng and Sui Tang 2022 *Inverse Problems* **38** 015004

View the [article online](#) for updates and enhancements.

You may also like

- [Parameter Estimation in Electrical Power Systems Using Prony's Method](#)
Claudius Ortbandt, Cezary Dzienis, Robert Matussek et al.

- [Performance Analysis of Optimal Analog Beamforming with Modified ESPRIT DOA Estimation Algorithm for MIMO NOMA System Over Rayleigh Fading Channel](#)
Yuvraj R and Aravind Kumar Y. V.

- [Quantum field tomography](#)
A Steffens, C A Riofrío, R Hübener et al.

Estimate the spectrum of affine dynamical systems from partial observations of a single trajectory data

Jiahui Cheng¹ and Sui Tang^{2,*} 

¹ Department of Mathematics, Georgia Institute of Technology, United States of America

² Department of Mathematics, University of California Santa Barbara, United States of America

E-mail: jcheng328@gatech.edu and suitang@ucsb.edu

Received 6 May 2021, revised 28 October 2021

Accepted for publication 9 November 2021

Published 1 December 2021



CrossMark

Abstract

In this paper, we study the nonlinear inverse problem of estimating the spectrum of a system matrix, that drives a finite-dimensional affine dynamical system, from partial observations of a single trajectory data. In the noiseless case, we prove an annihilating polynomial of the system matrix, whose roots are a subset of the spectrum, can be uniquely determined from data. We then study which eigenvalues of the system matrix can be recovered and derive various sufficient and necessary conditions to characterize the relationship between the recoverability of each eigenvalue and the observation locations. We propose various reconstruction algorithms with theoretical guarantees, generalizing the classical Prony method, ESPRIT, and matrix pencil method. We test the algorithms over a variety of examples with applications to graph signal processing, disease modeling and a real-human motion dataset. The numerical results validate our theoretical results and demonstrate the effectiveness of the proposed algorithms.

Keywords: sampling and reconstruction, Prony method, matrix pencil, ESPRIT method, affine linear system, spectrum estimation, partial observation

(Some figures may appear in colour only in the online journal)

1. Introduction

Many physical processes in science and engineering are modeled as linear dynamical systems with a state-space formulation. For example, linear time-invariant systems are widely used to characterize electrical systems and their properties [25]. Another common example is provided

*Author to whom any correspondence should be addressed.

by the diffusion processes over the graphs, which have found wide applications including modeling rumor propagation in social networks [51], traffic movement in transportation network [19], spatial temperature profiles over sensor networks [49], and neural activities at different regions of the brain [43]. In these applications, the states of the dynamical system at different time instances refer to *signals* of interest. In practice, a network of sensors is often placed to measure the values of evolving signals with varying locations, and the collected data are called *samples*. A fundamental inverse problem is to recover the dynamical system from *samples* of evolving signals.

In the case of known dynamics, the inverse problem reduces to the recovery of the initial state and is also called the source localization problem. But in many cases, the dynamics are also unknown and need to estimate from the data. Recently, this type of inverse problem has attracted a lot of attention in the graph signal processing community [20, 29, 30, 32–34, 41, 49]: the system matrices of the underlying dynamical systems are related to the topology of the underlying graph, such information is not available in many applications, and needs to be estimated. This is not only for enhancing data processing tasks but also for data interpretability, i.e., the graph topology provides an abstraction for the underlying data dependencies.

The previous methods typically assume the signals are fully observed. However, one may only afford to measure the values of the signals at a subset of coordinates, due to the high cost of building accurate sensors and application-specific restrictions. As a result, there is a possible significant loss of spatial information in each step of data acquisition. The inverse problem becomes in general ill-posed. In particular, in certain situations, we are only able to measure the dynamical system from a single trajectory, since the measurement process results in the destruction or alteration of the system under study. In such scenarios, the exact recovery of the dynamical systems is in general infeasible. However, one may still hope to recover the key information of the dynamical system.

The spectrum of the system matrix provides valuable information about the underlying dynamical system. For example, for linear autonomous systems, the spectrum plays a crucial role in analyzing the stability of the dynamical system. For diffusion processes over the graphs, the spectrum of the system matrix also implies the spectrum of the graph Laplacian matrix, which reveals the structure information of the underlying graphs. In the case of normalized graph Laplacian (see definition 6 and its remark), we can determine the number of components and infer whether the graph is complete or bipartite, see section 1.3, lemma 1.7 in [17]. The connection between the graph properties and the characteristic polynomial, eigenvalues, and eigenvectors of matrices associated with the graph, such as its adjacency matrix or other types of Laplacian matrix, has been extensively studied in the field of spectral graph theory, the reader can refer to [17, 42] for more details.

In this paper, we are interested in the recovery of the spectrum of the system matrix that generates the dynamical system from partial observations of a single trajectory. Affine dynamical systems are a natural starting point because they have a simple structure yet broad applications, including random walks on graphs [31], diffusion processes [49], linear mechanical and electronic systems [25], compartmental models in biological modeling such as pharmacodynamics, gene regulation [12, 23, 26]. The inverse problem in this setting is nonetheless nontrivial because the solution to such a system depends nonlinearly on the system matrix. We shall begin with a discrete finite-dimensional affine system:

$$x_{t+1} = Ax_t + c, t = 0, 1, 2, \dots, \quad (1)$$

$$x_0 = b. \quad (2)$$

In (1), $x_t \in \mathbb{C}^d$ is the state of the system at time t , the vector $b \in \mathbb{C}^d$ is the unknown initial state and the vector $c \in \mathbb{C}^d$ can be viewed as an unknown control or external force term. The system matrix $A \in \mathbb{C}^{d \times d}$ is unknown. We shall also consider the continuous-times analogue of (1):

$$\dot{x}(t) = Ax(t) + c, t \geq 0, \quad (3)$$

$$x(0) = b. \quad (4)$$

We then formulate the inverse problem in the most general setting, which we call the *dynamical sampling* problem, as follows:

Problem 1 (Dynamical sampling problem). *Suppose that we observe the affine dynamical system at time instances τ , for each $t \in \tau$, only part of the state x_t is observed, $\{x_t(i) : i \in \Omega_t\}$ where $\Omega_t \subset \{1, \dots, d\}$, under what conditions on observed locations $\{\Omega_t\}_{t \in \tau}$, the initial condition b and the control term c such that the key parameters of A can be recovered from the space–time samples? If so, what are algorithms that perform efficient reconstructions?*

Problem 1 exhibits features that are similar to many fundamental problems in the interface of signal processing, machine learning, and control theory of dynamical systems: observability of the dynamical systems, network topology identification, super-resolution, deconvolution, completion of the low-rank matrices. However, even in the most basic cases, the dynamical sampling problems are different and necessitate new theoretical and algorithmic techniques.

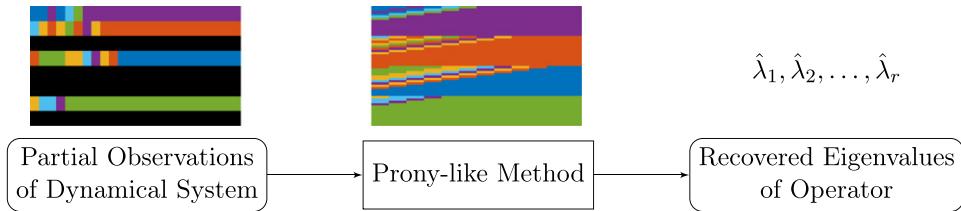
In this paper, we investigate the case where $\Omega_t \equiv \Omega \subset [d] := \{1, 2, \dots, d\}$ and restrict our attention to the recovery of eigenvalues of A . We define by S_Ω the observation matrix that

$$S_\Omega x_t = \sum_{i \in \Omega} x_t(i) e_i, \quad (5)$$

where $\{e_i\}_{i=1}^d$ is the standard orthonormal basis in \mathbb{C}^d . Given the partial observation of a single trajectory

$$\{S_\Omega x_t, t = 0, 1, \dots\}, \quad (6)$$

we develop theory and algorithms for solving the inverse problem of recovering eigenvalues of A .



1.1. Summary of contributions

We begin by showing that, an annihilating polynomial of A related to Ω and system parameters can be uniquely determined from (5), whose roots are a subset of eigenvalues of A . We then derive necessary and sufficient conditions on the interplay between Ω and the system parameters to characterize which eigenvalues of A are Ω -recoverable. In particular, we provide characterizations on the universal construction of Ω that allows the recovery of all eigenvalues

of A almost surely. Our theorems shed light on the minimal cardinality of Ω and guide constructions of Ω to recover the target eigenvalues. For numerical algorithms, our proofs provide a Prony-type method to reconstruct eigenvalues. We also generalize the classical matrix pencil method and estimation of signal parameters via rotational invariant techniques (ESPRIT) to recover eigenvalues and provide theoretical guarantees. Finally, we perform a systematic numerical study to examine the accuracy of the reconstruction and compare the performance of the proposed algorithms for various observational parameters on a variety of examples. We also test the effectiveness of our approach on a real human motion data set.

In summary, the main contributions of this paper are (i) characterizing the uniqueness of annihilating polynomial and the relationship between Ω and recoverable eigenvalues for affine systems; (ii) proposing various algorithms based on the Prony-method, matrix pencil, and ESPRIT with superior numerical performance.

Our work is build upon recent progress in studying dynamical sampling problems where A is known and the goal is to recover the initial state ([2, 7, 48], see [1, 3–5, 8, 14, 16] for recent developments) and the work on system identification aspect [6, 47]. In [6], Aldroubi *et al* studied the homogeneous case when $c = 0$ and A is diagonalizable. They propose Prony-type methods but no numerical examples were presented. In this work, we extend their analysis to affine systems with general system matrices and propose various algorithms. This extension allows broad applications, such as random walks and diffusion processes on directed graphs, and compartmental models in disease modeling. We further derive results for universal selections of observation locations and the continuous-time systems.

During the finalization of our work, we noticed the recent work [10] where they propose approximate Prony method to recover the eigenvalue of A in the homogeneous linear dynamical system when the data samples lose the sign information. They prove results on uniqueness for the cases where A is diagonalizable and the eigenvalues of A are collision-free. Our results hold for arbitrary matrix A and we allow the multiplicities of eigenvalues greater than 1. We would expect that our method can be extended to the case of the phaseless samples.

1.2. Connection with other fields/related work

Connection to system identification problem. Consider a time-invariant linear dynamical system.

$$x_{t+1} = Ax_t + Bc_t, \quad (7)$$

$$y_t = Cx_t, \quad (8)$$

where $c_t \in \mathbb{R}^d$ is the input vector and $y_t \in \mathbb{R}^s$ is the output vector. The parameter estimation problem considered in control theory aims to recover the parameter matrices $A, B \in \mathbb{R}^{d \times d}$, $C \in \mathbb{R}^{s \times d}$ from the output vectors $\{y_t\}$, provided a set of input vectors $\{x_t\}$. The classical results show that C has to be full-rank (i.e., the states are all observed) to make identifiability of parameter matrices possible, see the survey in [50] as well as the introduction section in [18]. In the special case of $B = 0$, and $C = I$, i.e., $\Omega_t \equiv \{1, \dots, d\}$ in our setting, the identifiability of A from a single trajectory with a fixed initial condition has been studied in [44] and later it has been studied in affine dynamical system ($B = I$, $C = I$, $c_t \equiv c$) in [21]. It has been shown that sufficient and necessary condition is that $\{x_0, Ax_0, \dots, A^L x_0\}$ needs to span \mathbb{C}^d , which means that A has only one Jordan block for each of its eigenvalues and certain constraints are imposed on x_0 , we refer to [21, 44] for more details. This tells us why in the case of partial observations from single trajectory, one should not hope for the full reconstruction of matrix A in general. It is also mentioned in [21] that one can predict the time-dependence for any observable component of the state variable, as long as a Hankel matrix constructed from data is invertible. This

claim now becomes a special case of proposition 2.1 developed in this paper. Our result further characterizes when this Hankel matrix is invertible.

Connection to structured signal recovery problem. Parameter estimation problems of structured signals have been extensively studied in signal processing. In [36], the authors present an abstract formulation: let V be a normed vector space over \mathbb{C} and A be a *known* linear operator from V to V , one is interested in recovering a signal $b \in V$ that is M -sparse with respect to eigenfunctions of A

$$b = \sum_{j \in J} c_j v_j, \quad \text{with } |J| = M. \quad (9)$$

The goal is to recover $\{c_j\}$ and $\{v_j\}$ from data samples $\mathcal{F}(A^l b)$ for $l = 0, 1, \dots, L$ where $\mathcal{F} : V \rightarrow \mathbb{C}$ is a linear functional. The specific instances include the super-resolution, blind deconvolution, recovery of signals with finite rate innovations, and we refer to [36] for more details. The keys behind successful recovery of the structured signal b are (1) the eigenvalues $\{\lambda_j\}_{j \in J}$ of A corresponding to $\{v_j\}_{j \in J}$ can be recovered from data under suitable assumptions on \mathcal{F} (2) A is known and its eigenvalues are assumed to have geometric multiplicity 1, therefore finding eigenvectors suffices to find their corresponding eigenvalues. In [36], the authors proposed generalized Prony method to recover the eigenvalues for $L = 2M - 1$.

Back to our setting, assume that A is diagonalizable and $c = 0$ in (1), and we take $\Omega_l \equiv \{e_i\}$ for some $i \in [d]$ so that the observational functional $\mathcal{F}(x_l) = \langle x_l, e_i \rangle = x_l(i)$. In this case, we use eigenvectors of A as basis and represent the initial condition b as in (9). Then the structured signal recovery problem becomes a special case of the dynamical sampling problem considered here. The theorem 1 developed in this paper generalizes the main results Theorems 2.1 and 2.3 in [36] in two aspects: (i) our result implies the sufficient conditions on \mathcal{F} is also necessary (ii) our result provided various sufficient and necessary characterizations to answer which eigenvalues of A are Ω -recoverable for arbitrary $\Omega \subset \{1, 2, \dots, d\}$.

1.3. Notation

In the following, we use standard notations. By \mathbb{N} , we denote the set of all positive integers. For a positive integer d , we use the notation $[d]$ to represent the set $\{1, 2, \dots, d\}$. The linear space of all column vectors with M complex components is denoted by \mathbb{C}^M . The linear space of all complex $M \times N$ matrices is denoted by $\mathbb{C}^{M \times N}$. For a matrix $A = (a_{ij}) \in \mathbb{C}^{M \times N}$, its transpose is denoted by A^T , its conjugate-transpose by A^* and its Moore–Penrose pseudoinverse by A^+ . For a vector $z = (z_i) \in \mathbb{C}^M$, the $M \times M$ diagonal matrix built from z is denoted by $\text{diag}(z)$. Further, we use submatrix notation similar to that of MATLAB. For example, if $A \in \mathbb{C}^{M,M+1}$, then $A(1:M, 2:M+1)$ is the submatrix of A obtained by extracting rows 1 through M and columns 2 through $M+1$, and $A(1:M, M+1)$ means the last column vector of A .

1.4. Preliminaries

Throughout the paper, we assume that the system matrix $A \in \mathbb{C}^{d \times d}$ and it has distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Consider the Jordan decomposition $A = U J U^{-1}$, where $U \in \mathbb{C}^{d \times d}$ is invertible and the Jordan matrix $J \in \mathbb{C}^{d \times d}$ is block diagonal defined as follows:

$$J = \begin{pmatrix} J_1 & O & \dots & O \\ O & J_2 & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & J_n \end{pmatrix}. \quad (10)$$

In (10), for $s = 1, \dots, n$, the Jordan block J_s corresponds to λ_s and $J_s = \lambda_s I_s + N_s$ where I_s is the identity matrix of dimension h_s , and N_s is a nilpotent block-matrix of dimension h_s :

$$N_s = \begin{pmatrix} N_{s_1} & 0 & 0 & 0 \\ 0 & N_{s_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & N_{s_{r_s}} \end{pmatrix}, \quad (11)$$

where each N_{s_i} is a $t_i^{(s)} \times t_i^{(s)}$ cyclic nilpotent matrix of the form (12),

$$N_{s_i} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ I_{t_i^{(s)}-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}, \quad (12)$$

with $t_1^{(s)} \geq t_2^{(s)} \geq \dots \geq t_{r_s}^{(s)}$ and $t_1^{(s)} + t_2^{(s)} + \dots + t_{r_s}^{(s)} = h_s$. Also $h_1 + \dots + h_n = d$.

Each Jordan block J_s corresponds to an invariant subspace V_s of A and the Jordan form J gives a decomposition of \mathbb{C}^d into invariant subspaces of A : $\mathbb{C}^d = \bigoplus_{s=1}^n V_s$. We define the projection onto V_s by $P(\lambda_s; A)$. Similarly, we denote by E_s the invariant subspace of J corresponding to the block J_s ; E_s is spanned by the canonical basis $\{e_i : i = h_1 + \dots + h_{s-1} + 1, \dots, h_1 + \dots + h_s\}$ ($h_0 = 0$ for the cases $s = 1$); the projection onto E_s is denoted by $P(\lambda_s; J)$ and we have $P(\lambda_s; J) = UP(\lambda_s; A)U^{-1}$. Any vector $f \in \mathbb{C}^d$ admits the unique decomposition $f = \sum_{s=1}^n f_s$, where $f_s = P(\lambda_s; J)f \in E_s$.

Note that $J|_{E_s} = \lambda_s I_s + N_s$, we will use $Jf_s = (\lambda_s I_s + N_s)f_s$ by viewing the vector $f_s \in \mathbb{C}^{h_s}$ and interpreting the vector $(\lambda_s I_s + N_s)f_s$ as its canonical embedding in \mathbb{C}^d .

With the abuse of notation, if $N_s \equiv 0$ for $s = 1, \dots, n$, then J in (10) reduced to a diagonal matrix of the form

$$D = \begin{pmatrix} D_1 & O & \dots & O \\ O & D_2 & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & D_n \end{pmatrix}. \quad (13)$$

In this case, A is diagonalizable.

Definition 1. Let k_i^s denote the row index corresponding to the entry 1 in the last nonzero column of the block N_{s_i} (12) from the matrix J (10), and let $e_{k_i^s}$ be the corresponding elements of the standard basis of \mathbb{C}^n , so that each $e_{k_i^s}$ is the cyclic vector associated to $N_{s_i}^*$. We also define $W_s = \text{span}\{e_{k_i^s} : i = 1, \dots, r_s\}$, for $s = 1, \dots, n$, and P_s will denote the orthogonal projection onto W_s . The family $P_J = \{P_j : j = 1 \dots n\}$ comprised of these projections will be called the *penthouse family* of the matrix J .

Let us give an example to illustrate the definition above. Consider a Jordan matrix as

$$J = \begin{pmatrix} J_1 & \\ & J_2 \end{pmatrix} = \begin{pmatrix} 3 & & & & \\ 1 & 3 & & & \\ & 1 & 3 & & \\ & & & 3 & \\ & & & 1 & 3 \\ & & & & 2 \end{pmatrix}. \quad (14)$$

It has three nilpotent blocks given as N_{11} , N_{12} and N_{21} . The cyclic vectors associated to N_{11}^* , N_{12}^* and N_{21}^* are e_3 , e_5 and e_6 , respectively.

Definition 2. We introduce three kinds of minimal polynomials that are useful in the paper.

- For $A \in \mathbb{C}^{d \times d}$, the minimal polynomial q^A is the monic polynomial of the smallest degree, such that $q^A(A) \equiv 0$, and we denote its degree as $r^A = \deg(q^A)$.
- For any matrix $S \in \mathbb{C}^{m \times d}$, the S -altered minimal polynomial of A , denoted by q_S^A , is the monic polynomial of smallest degree among all the polynomials p such that $Sp(A) = 0$ and $r_S^A := \deg(q_S^A)$
- The A -minimal polynomial q_b^A for a vector b in \mathbb{C}^d is the monic polynomial of the smallest degree, such that $q_b^A(A)b \equiv 0$, and we denote its degree as $r_b^A = \deg(q_b^A)$.

Definition 3. A Krylov space of order r generated by $A \in \mathbb{C}^{d \times d}$ and $b \in \mathbb{C}^d$, is defined by

$$\mathcal{K}_r(A, b) := \text{span}\{b, Ab, \dots, A^{r-1}b\}. \quad (15)$$

In particular, for any $r \geq r_b^A - 1$, we will denote $\mathcal{K}_r(A, b) = \mathcal{K}_\infty(A, b)$.

Definition 4. Let $S \in \mathbb{C}^{m \times d}$, $A \in \mathbb{C}^{d \times d}$ and $f \in \mathbb{C}^d$. The (S, A, b) -annihilating polynomials are all polynomials q such that $Sq(A)\mathcal{K}_S^A(A, b) = \{0\}$. We denote by $r_{S,b}^A$, the smallest degree among all monic (S, A, b) -annihilating polynomials.

1.5. Solution formulas to affine systems

In this section, we present the explicit solution formulas in terms of A, b and c for system (1) and its continuous-time counterpart, which will be useful in our analysis. Let x_t be the solution of discrete system (1). Then using the recursive relation, we obtain that

$$x_t = A^t b + (A^{t-1} + \dots + I)c, \quad t = 1, 2, \dots \quad (16)$$

For the continuous-time system (3), we let x_t^{cont} be the solution of system (3). Then the solution can be obtained by differentiating (3) and solving corresponding linear initial value problem for \dot{x}_t :

$$x_t^{\text{cont}} = e^{tA}b + g(t; A)c, \quad t \geq 0, \quad (17)$$

where

$$g(t; A) = \sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+1)!} A^k = tI + \frac{t^2}{2}A + \frac{t^3}{3!}A^2 + \dots \quad (18)$$

We also list several useful properties of $g(t; A)$:

Lemma 1.

$$\frac{d}{dt}g(t; A) = e^{tA}; \quad Ag(1; A) = g(1; A)A = e^A - I.$$

In particular, if A is invertible, then $g(1; A) = (e^A - I)A^{-1}$.

Proof. Note that the derivative of each term occurring in equation (18) is

$$\frac{d}{dt} \left(\frac{t^{k+1}}{(k+1)!} A^k \right) = \frac{t^k}{k!} A^k, \quad k = 0, 1, 2, \dots$$

Then from the sum rule for the derivative, we have

$$\begin{aligned} \frac{d}{dt}g(t; A) &= \frac{d}{dt} \left(\sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+1)!} A^k \right) \\ &= \sum_{k=0}^{\infty} \left[\frac{d}{dt} \left(\frac{t^{k+1}}{(k+1)!} A^k \right) \right] \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = e^{tA}. \end{aligned}$$

Also, plugging $t = 1$ into equation (18) to get $g(1; A)$, we obtain that

$$\begin{aligned} Ag(1; A) &= A \left(\sum_{k=0}^{\infty} \frac{1}{(k+1)!} A^k \right) \\ &= \sum_{k=0}^{\infty} \frac{1}{(k+1)!} A^{k+1} \\ &= e^A - I, \end{aligned}$$

where the last equality also comes from the power series expansion of exponential function e^A . \square

2. Main results

2.1. Discrete-time affine dynamical systems

2.1.1. Case $c = 0$. We will begin with the homogeneous linear dynamical system with $c = 0$. In this case, the discrete dynamical system reduces to

$$x_t = A^t x_0, \quad x_0 = b. \quad (19)$$

Given the observation locations $\Omega \subset [d]$, we show that the minimal (S_Ω, A, b) -annihilating polynomial $q_{S_\Omega, b}^A$ can be uniquely recovered from $\{S_\Omega(x_t) : t = 0, 1, \dots\}$ and the roots of $q_{S_\Omega, b}^A$ are eigenvalues of A . Here by uniqueness, we mean that if $S_\Omega x_t = S_\Omega \tilde{x}_t$ for $t = 0, 1, \dots$,

where \tilde{x}_t is the trajectory of the system (19) with a system matrix \tilde{A} and an initial state \tilde{b} , then $q_{S_\Omega, b}^A = q_{S_\Omega, \tilde{b}}^{\tilde{A}}$.

Proposition 1. *Given partial observations of trajectory data determined by $\Omega \subset [d]$*

$$S_\Omega(x_t), \quad t = 0, \dots, 2r-1, \quad r = r_{S_\Omega, b}^A,$$

we construct a Hankel matrix

$$H := \begin{bmatrix} S_\Omega b & \dots & S_\Omega A^{r-1}b \\ \vdots & \dots & \vdots \\ S_\Omega A^{r-1}b & \dots & S_\Omega A^{2r-2}b \end{bmatrix}. \quad (20)$$

Then H is of full column rank and there exist a unique solution $q = [q_1, \dots, q_r]^T \in \mathbb{C}^r$ to the linear system

$$Hq = -h_{\Omega, r}, \quad h_{\Omega, r} = \begin{bmatrix} S_\Omega A^r b \\ \vdots \\ S_\Omega A^{2r-1} b \end{bmatrix}. \quad (21)$$

We have that, $q_{S_\Omega, b}^A(z) = z^r + \sum_{i=1}^r q_i z^{i-1}$. In addition, the roots of $q_{S_\Omega, b}^A$ are eigenvalues of A .

Proof. Suppose that $Hq = -h_r$. Let the polynomial $q(z) = z^r + \sum_{k=0}^{r-1} q_{k+1} z^k$. Then it follows that

$$S_\Omega q(A) A^t b = 0, \quad t = 0, \dots, r-1.$$

By lemma 5, the solution is unique and $q = q_{S_\Omega, b}^A$. It follows from the equation (57) in lemma 4 (see appendix), that the roots of $q_{S_\Omega, b}^A$ are eigenvalues of A . \square

Proposition 1 in fact provides us with a Prony-type algorithm to reconstruct the annihilating polynomial $q_{S_\Omega, b}^A$ from data, by solving the Hankel-type equation (21). Then we find roots of $q_{S_\Omega, b}^A$ which are a part of eigenvalues of A . However, a key question still not addressed is which eigenvalues of A can be recovered. Our goal is to find the relationship between Ω and recoverable eigenvalues. Such results are useful in the selection of Ω to recover the target eigenvalues.

Theorem 1. *Assume that the evolution matrix $A \in \mathbb{C}^{d \times d}$ and its Jordan decomposition can be written as $A = UJU^{-1}$ where J is a Jordan matrix as in (10). Let $b \in \mathbb{C}^d$ be the initial state. Then the polynomial $q_{S_\Omega, b}^A$ can be uniquely determined from $\{S_\Omega x_t : t = 0, 1, \dots\}$. In addition, the following statements are equivalent: for $s = 1, \dots, n$,*

- (a) λ_s is **not** a root of $q_{S_\Omega, b}^A$.
- (b) $(U^{-1}b)_s \perp \mathcal{K}_\infty(N_s^*, (U^*e_i)_s)$ for all $i \in \Omega$, where $(U^{-1}b)_s = P(\lambda_s; J)U^{-1}b$, $(U^*e_i)_s = P(\lambda_s; J)U^*e_i$.
- (c) $P^*(\lambda_s; A)e_i \perp \mathcal{K}_\infty(A, b)$ for all $i \in \Omega$, where $P^*(\lambda_s; A)$ denotes the adjoint operator of $P(\lambda_s; A)$.

Proof. The proof is based on lemmas 3–5 in the appendix.

- Claim 1: the polynomial $q_{S_\Omega, b}^A$ has λ_s as one of its roots if and only if $\deg(q_{S_\Omega, (U^{-1}b)_s}^J) := r_{S_\Omega, (U^{-1}b)_s}^J \geq 1$.

On one hand, $q_{S_\Omega, b}^A = q_{S_\Omega, U^{-1}b}^J$. On the other hand, by lemma 4, we have $q_{S_\Omega, U, U^{-1}b}^J = \prod_{s=1}^n q_{S_\Omega, (U^{-1}b)_s}^J$ where $(U^{-1}b)_s = P(\lambda_s; J)U^{-1}b$ and $q_{S_\Omega, (U^{-1}b)_s}^J(z) = (z - \lambda_s)^{r_{S_\Omega, (U^{-1}b)_s}^J}$. Therefore claim 1 is proved.

- Claim 2: $r_{S_\Omega, (U^{-1}b)_s}^J = 0$ if and only if

$$(U^{-1}b)_s \perp \text{span}\{N_s^{*t} \hat{u}_{s,i}, t = 0, 1, \dots\} = \text{span}\{N_s^{*t} \hat{u}_{s,i}, t = 0, 1, \dots, m_{s,i} - 1\}, i \in \Omega,$$

where $\hat{u}_{s,i} = (U^* e_i)_s$, and $m_{s,i} = r_{\hat{u}_{s,i}}^{N_s^*}$.

By definition 4, $q_{S_\Omega, (U^{-1}b)_s}^J$ satisfies

$$S_\Omega U q_{S_\Omega, (U^{-1}b)_s}^J(J) J^t (U^{-1}b)_s = 0, \quad t = 0, 1, \dots, r_{(U^{-1}b)_s}^J - 1. \quad (22)$$

Note that $(U^{-1}b)_s \in E_s := P(\lambda_s; J)$ and therefore we can restrict J on E_s , i.e., use $J|_{E_s} = \lambda_s I_s + N_s$ instead of J . By the proof of lemma 4, we have $q_{S_\Omega, (U^{-1}b)_s}^J(J|_{E_s}) = N_s^{r_{S_\Omega, (U^{-1}b)_s}^J}$, and the equations in (22) are equivalent to

$$S_\Omega U N_s^{r_{S_\Omega, (U^{-1}b)_s}^J} (\lambda_s I_s + N_s)^t (U^{-1}b)_s = 0, \quad t = 0, 1, \dots, r_{(U^{-1}b)_s}^J - 1, \quad (23)$$

which can be further simplified as

$$S_\Omega U N_s^{r_{S_\Omega, (U^{-1}b)_s}^J} N_s^t (U^{-1}b)_s = 0, \quad t = 0, 1, \dots, r_{(U^{-1}b)_s}^J - 1. \quad (24)$$

Representing each equation in (24) using the inner product, we obtain that, for $i \in \Omega$

$$\hat{u}_{s,i} \perp \text{span}\{N_s^t (U^{-1}b)_s, t = 0, 1, \dots, r_{(U^{-1}b)_s}^J - 1\}, \quad (25)$$

where $\hat{u}_{s,i} = P(\lambda_s; J)U^* e_i$.

Note that $\text{span}\{N_s^t (U^{-1}b)_s, t = 0, 1, \dots, r_{(U^{-1}b)_s}^J - 1\} = \text{span}\{N_s^t (U^{-1}b)_s, t = 0, 1, \dots\}$, and $\langle \hat{u}_{s,i}, N_s^t (U^{-1}b)_s \rangle = \langle N_s^{*t} \hat{u}_{s,i}, (U^{-1}b)_s \rangle$, claim 2 is followed by the fact that

$$\text{span}\{N_s^{*t} \hat{u}_{s,i}, t = 0, 1, \dots\} = \text{span}\{N_s^{*t} \hat{u}_{s,i}, t = 0, 1, m_{s,i} - 1\},$$

where $m_{s,i} = r_{\hat{u}_{s,i}}^{N_s^*}$ is the least integer m such that $N_s^{*m-1} (U^{-1}b)_s \neq 0$ but $N_s^{*m} (U^{-1}b)_s = 0$.

- Claim 3: the equations in (22) are equivalent to

$$\langle P^*(\lambda_s; A) e_i, A^t b \rangle = 0, \quad t = 0, 1, \dots, i \in \Omega.$$

Note that

$$\begin{aligned}
\langle \hat{u}_{s,i}, J_s^t(U^{-1}b)_s \rangle &= \langle P(\lambda_s; J)U^*e_i, J_s^t(U^{-1}b)_s \rangle \\
&= \langle e_i, UP(\lambda_s; J)J_s^t(U^{-1}b)_s \rangle \\
&= \langle e_i, UP(\lambda_s; J)J^tP(\lambda_s; J)U^{-1}b \rangle \\
&= \langle e_i, UP(\lambda_s; J)J^tU^{-1}b \rangle \\
&= \langle e_i, UP(\lambda_s; J)U^{-1}UJ^tU^{-1}b \rangle \\
&= \langle e_i, P(\lambda_s; A)A^t b \rangle = \langle P^*(\lambda_s; A)e_i, A^t b \rangle,
\end{aligned}$$

where we used the facts that $P(\lambda_s; J)^2 = P(\lambda_s; J)$, $P(\lambda_s; J)J = JP(\lambda_s; J)$ and $P(\lambda_s; A) = UP(\lambda_s; J)U^{-1}$. Then the conclusion follows in a similar way as in the proof of claim 2. \square

In theorem 1, we view Ω as a set of functionals $\{e_i : i \in \Omega\}$ in the dual space. Part (b) characterizes the recoverability of the eigenvalue λ_s by checking the local orthogonality of the vector $(U^{-1}b)_s$ with the Krylov subspace generated by N_s^* and $\{(U^*e_i)_s : i \in \Omega\}$. Part (c) provides an equivalent geometric characterization on the global Krylov subspace generated by the trajectory of the system and the projection of the vectors $\{U^*e_i : i \in \Omega\}$ onto the invariant subspace E_s . In this way, we find the interplay between A, b and Ω that allows the recovery of the eigenvalue λ_s .

Remark 1. Theorem 1 also holds for a diagonalizable matrix A . Suppose that $A = UDU^{-1}$ where D is a diagonal matrix as in (13). By setting $N_s = 0$, the part (b) in this case can be simplified as

$$(U^{-1}b)_s \perp (U^*e_i)_s,$$

for all $i \in \Omega$, where $(U^{-1}b)_s = P(\lambda_s; D)U^{-1}b$, and $(U^*e_i)_s = P(\lambda_s; D)U^*e_i$. This condition was proved in theorem 3.7 in [6]. In particular, when $P(\lambda_s; D)$ is a rank one projection, λ_s is recoverable if and only if $(U^{-1}b)_s(U^*e_i)_s \neq 0$ for some $i \in \Omega$, which was listed as a sufficient condition in theorem 2.1 in [36] for $\Omega = \{e_i\}$.

Universal constructions of Ω . From theorem 1, we know that if $(U^{-1}b)_s = P(\lambda_s; J)U^{-1}b = 0$, then λ_s can not be recovered no matter what choice of Ω is. However, the set consisting of this kind of initial conditions is of measure zero. Let us consider a generic set of \mathbb{C}^d defined by $\mathcal{S} = \{b \in \mathbb{C}^d : (U^{-1}b)_s \neq 0 \text{ for } s = 1, \dots, n\}$. If the initial state b is sampled from a non-degenerate probability measure on \mathbb{C}^d , then $b \in \mathcal{S}$ almost surely. One may ask: how to choose Ω such that all eigenvalues of A can be recovered from partial observations of a single trajectory starting from $b_0 \in \mathcal{S}$. This question yields the following definition.

Definition 5 (Universal sampling set). A set Ω is said to be universal for the system (19) if all eigenvalues of A can be recovered from the partial observation of the trajectory data $S_\Omega x_t$ for $t = 0, 1, \dots, 2r_{S_\Omega, b}^A - 1$ with $x_0 = b \in \mathcal{S}$.

Based on theorem 1, we derive various characterizations on universal sampling sets Ω that can guide their constructions.

Theorem 2. *Assume that the evolution matrix $A \in \mathbb{C}^{d \times d}$ and its Jordan decomposition can be written as $A = UJU^{-1}$ where J is a Jordan matrix as in (10). The following statements are equivalent*

- (a) Ω is universal for the system (19).
- (b) The set of vectors $\{e_i, A^*e_i, \dots, (A^*)^{r_{e_i}^A-1}e_i : i \in \Omega\}$ span \mathbb{C}^d . In other words, $\sum_{i \in \Omega} \mathcal{K}_\infty(A^*; e_i) = \mathbb{C}^d$.
- (c) The set of vectors $\{P_s U^* e_i, i \in \Omega\}$ spans $\text{Range}(P_s)$ for $s = 1, \dots, n$, where $P_J = \{P_s : s = 1, \dots, n\}$ is the penthouse family for J introduced in definition 1, in other words, P_s is the orthogonal projection onto the span of cyclic vectors corresponding to the nilpotent block N_s^* .

Proof. Note that $A^* = (U^*)^{-1}J^*U^*$, we have that

$$\sum_{i \in \Omega} \mathcal{K}_\infty(A^*; e_i) = \mathbb{C}^d \iff \sum_{i \in \Omega} \mathcal{K}_\infty(J^*; U^*e_i) = \mathbb{C}^d,$$

which is also equivalent to

$$\sum_{i \in \Omega} P(\lambda_s^*; J^*) \mathcal{K}_\infty(J^*; U^*e_i) = E_s := \text{Range}(P(\lambda_s^*; J^*)), \quad s = 1, \dots, n.$$

The first equivalence is due to the invertibility of U and the second equivalence is due to the facts that $\sum_{s=1}^n P(\lambda_s^*; J^*) = I$ and the projections are mutual orthogonal.

The equivalence between (a) and (b) can then be proved using part (c) of theorem 1. Note that

$$\begin{aligned} \langle P^*(\lambda_s; A)e_i, A^t b \rangle &= \langle P(\lambda_s^*; A^*)e_i, A^t b \rangle \\ &= \langle P(\lambda_s^*; A^*)(A^*)^t e_i, b \rangle \\ &= \langle (U^*)^{-1} P(\lambda_s^*; J^*) U^* (U^*)^{-1} (J^*)^t U^* e_i, b \rangle \\ &= \langle P(\lambda_s^*; J^*) (J^*)^t U^* e_i, U^{-1} b \rangle \\ &= \langle P(\lambda_s^*; J^*) (J^*)^t U^* e_i, P(\lambda_s^*; J^*) U^{-1} b \rangle, \end{aligned} \quad (26)$$

where we used the facts that $P(\lambda_s^*; A^*) = P^*(\lambda_s; A)$, $P(\lambda_s^*; A^*)A^* = A^*P(\lambda_s^*; A^*)$, and $P(\lambda_s^*; J^*) = P(\lambda_s; J)$.

(b) \implies (a): for any $b \in \mathcal{S}$, and any $s \in [n]$, we have that $P(\lambda_s; J)U^{-1}b = P(\lambda_s^*; J^*)U^{-1}b \neq 0$. From (26), we see that it is impossible to have

$$\langle P^*(\lambda_s; A)e_i, A^t b \rangle = 0 \quad \text{for all } i \in \Omega,$$

since $\sum_{i \in \Omega} P(\lambda_s^*; J^*) \mathcal{K}_\infty(J^*; U^*e_i) = E_s$ and $P(\lambda_s^*; J^*)U^{-1}b$ is a nonzero vector in E_s . According to part (c) of theorem 1, λ_s is a root of $q_{S_\Omega, b}^A$. Therefore, all eigenvalues of A can be recovered.

(a) \implies (b): if there exists an $s \in [n]$, such that $\sum_{i \in \Omega} P(\lambda_s^*; J^*) \mathcal{K}_\infty(J^*; U^* e_i)$ is a proper subspace of E_s . Let the nonzero vector $f_s \in E_s$ be orthogonal to $\sum_{i \in \Omega} P(\lambda_s^*; J^*) \mathcal{K}_\infty(J^*; U^* e_i)$. Then for any $b \in \mathcal{S}$, the vector $\tilde{b} = b - UP(\lambda_s; J)b + Uf_s \in \mathcal{S}$, and we have

$$\langle P^*(\lambda_s; A)e_i, A^t \tilde{b} \rangle = 0 \quad \text{for all } i \in \Omega,$$

which yields a contradiction for the universality of Ω .

Now we prove the equivalence of (b) and (c). We use theorem 6 in the [appendix](#), which is a slight modification of the theorem 2.6 in [2]. \square

Part (b) of theorem 2 relates the universal construction of Ω to the trajectory behavior of a conjugate dynamical system to (19): consider trajectories starting with initial conditions $\{e_i : i \in \Omega\}$, they should not belong to any proper subspace. Part (c) of theorem 2 indicates that the universal construction only depends on the invertibility of certain submatrices of U . It guides us to construct universal sampling sets according to the spectral structure of A . For simplicity, we present an example for a diagonalizable matrix.

If $A = UDU^{-1}$ where D is a diagonal matrix as in (13), then the part (c) can be simplified as $\{P(\lambda_s; D)U^* e_i, i \in \Omega\}$ spans $\text{Range}(P(\lambda_s; D)) = E_s$. That is to say, the submatrix of U^* formed by selecting column indices according to Ω and row indices according to E_s needs to have full row rank. If we know $\dim(E_s) \equiv 1$ for $s = 1, \dots, d$, then it suffices to find a column of U^* which is nonzero everywhere. This example can be immediate to generalize to the Jordan case where there is only one cyclic vector for each nilpotent block. Below, we present a more complicated example.

Example 2.1 (Diffusion process over circulant graphs). In this system, the dynamics is generated by a circulant matrix A . It is well-known that A admits the spectral decomposition $A = UDU^{-1}$, where U is the discrete Fourier matrix and D is a diagonal matrix, with eigenvalues $\lambda_1, \dots, \lambda_n$. Note that in this case, D is similar to the standard diagonal form in (13) up to a permutation of the diagonal entries. Let $E_s := \text{Range}(P(\lambda_s; D))$ and $s_{\max} := \max_{s=1, \dots, n} \dim(E_s)$. Let us choose $\Omega = \{1, \dots, s_{\max}\}$, and denote by $(U^*)_\Omega$ the submatrix of U^* formed by choosing columns according to Ω . Then the matrix $(U^*)_\Omega$ satisfies the full spark property: any s_{\max} rows of $(U^*)_\Omega$ will form a Vandermonde matrix with distinct nodes, and therefore is invertible. As a result, $\{P(\lambda_s; D)U^* e_i, i \in \Omega\}$ spans E_s for $s = 1, \dots, n$ and Ω is a universal construction.

Example 2.1 reveals a connection of universal sampling sets with full-spark submatrices of U^* , and the latter problem has been extensively studied in the field of compressed sensing. For discrete Fourier matrix, Chebotarëv showed that every square submatrix of the discrete Fourier matrices is invertible if the dimension d is prime [46]. In this case, an arbitrary set $\Omega \subset [d]$ with $|\Omega| = s_{\max}$ is universal. In general, deterministic constructions are very difficult. As is done in compressed sensing, one can look for the random constructions of Ω at uniform such that any s_{\max} rows of $(U^*)_\Omega$ satisfies the restricted isometry property. It is shown that the cardinality of Ω will be slightly more than s_{\max} , with proportional factor mainly depending on the coherence of the columns U^* , a quantity measuring the dependence. The reader can refer to [15, 48] for more details.

2.1.2. Case $c \neq 0$. In this subsection, we deal with the affine systems with $c \neq 0$. Inspired by the idea in [21], we show that the affine systems can be transformed to a homogeneous system by a linear transformation. In this way, we can extend the previous results to the affine systems. For the sake of conciseness, we only state the generalization of theorem 1. Other results in section 2.1.1 can be derived similarly.

Now let us consider the affine system:

$$x_{t+1} = Ax_t + c, \quad x_0 = b, \quad t = 0, 1, \dots$$

Using the solution formula (16), we have $x_{t+1} - x_t = A^t((A - I)b + c)$. Now we denote $y_t = x_{t+1} - x_t$, then we have.

$$y_t = A^t y_0, \quad t = 0, 1, 2, \dots, \quad (27)$$

$$y_0 = (A - I)b + c. \quad (28)$$

Now we apply theorem 1 to the homogeneous system (27), we obtain the following theorem:

Theorem 3. *Assume that the evolution matrix $A \in \mathbb{C}^{d \times d}$ and its Jordan decomposition can be written as $A = UJU^{-1}$ where J is a Jordan matrix as in (10). Let $b \in \mathbb{C}^d$ be the initial state and c be the force term of the system. Define the vector $w = (A - I)b + c$. Then the polynomial $q_{S_\Omega, w}^A$ can be uniquely determined from $\{S_\Omega x_t : t = 0, 1, \dots\}$. In addition, the following statements are equivalent: for each $s \in [n]$*

- (a) λ_s is **not** a root of $q_{S_\Omega, w}^A$.
- (b) $w_s \perp \mathcal{K}_\infty(N_s^*, \hat{u}_{s,i})$ for all $i \in \Omega$ where $w_s = P(\lambda_s; J)U^{-1}w$, $\hat{u}_{s,i} = P(\lambda_s; J)U^*e_i$.
- (c) $P^*(\lambda_s; A)e_i \perp \mathcal{K}_\infty(A; w)$ for all $i \in \Omega$.

Now we compare theorem 3 with theorem 1. The only difference is to replace the vector b in theorem 1 with w . Note that $w = (A - I)b + c$, so the nonzero external vector c changes the geometry of vectors and Krylov spaces. As a result, for $\Omega \subset [d]$, λ_s could be recovered in the affine system but may not in its homogeneous counterpart. In particular, when A is the identity matrix, the initial condition b plays no role in determining the recoverability of the eigenvalues in affine systems. This can also be seen from the solutions to the systems: the initial condition b just adds a simple translation on the solutions. However, universal sampling sets for both affine system and its homogeneous counterpart are the same, since they only depend on the spectral matrix U by theorem 2.

2.2. Extension to continuous-time affine dynamical systems

In this section, we consider the continuous-time systems:

$$\dot{x}(t) = Ax(t) + c, \quad t \geq 0,$$

$$x(0) = b.$$

Suppose that observation time instances are uniform, without loss of generality, say at $t = 0, 1, 2, \dots$, we define $B = e^A$, and introduce $y_t = x_{t+1}^{\text{cont}} - x_t^{\text{cont}}$, then we have

$$y_t = B^t y_0, \quad (29)$$

$$y_0 = (B - I)b + g(1; A)c, \quad (30)$$

where we used the properties

$$g(t; A) = tg(1; tA), \quad (31)$$

$$(t+1)g(1; (t+1)A) - tg(1; tA) = e^{tA}g(1; tA). \quad (32)$$

Therefore, we can apply the results of discrete-time systems to system (29) generated by the evolution matrix B . In general, the mapping $A \rightarrow e^A$ does not preserve the Jordan decomposition of A except for the case when A is similar to a diagonal matrix. For easy presentation, we state the result for the diagonalizable case.

Theorem 4. *Assume that the evolution matrix $A = UDU^{-1}$ where D is a diagonal matrix. Let $b \in \mathbb{C}^d$ be the initial state and c be force term of the system. Define the matrix $B = e^A$ and the vector $w = (B - I)b + g(1; A)c$. Then the polynomial $q_{S_\Omega, w}^B$ can be uniquely determined from $\{S_\Omega x_t^{\text{cont}} : t = 0, 1, \dots\}$. In addition, the following statements are equivalent: for each $s \in [n]$*

- (a) e^{λ_s} is **not** a root of $q_{S_\Omega, w}^B$.
- (b) $(U^{-1}w)_s \perp \text{span}\{\hat{u}_{s,i}, i \in \Omega\}$, where the vector

$$(U^{-1}w)_s = \begin{cases} (e^{\lambda_s} - 1)P(\lambda_s; D)U^{-1} \left(b + \frac{1}{\lambda_s}c \right), & \text{if } \lambda_s \neq 0 \\ P(\lambda_s; D)U^{-1}c, & \text{otherwise,} \end{cases}$$

and the vector $\hat{u}_{s,i} = P(\lambda_s; D)U^*e_i$.

- (c) $P^*(\lambda_s; B)e_i \perp \mathcal{K}_\infty(B; w)$ for all $i \in \Omega$.

Proof. Since $A = UDU^{-1}$, we have that $B = Ue^D U^{-1}$. That is, the spectrum of B consists of $e^{\lambda_1}, \dots, e^{\lambda_s}$. We first show the formula of $(U^{-1}w)_s$ presented in part (b). Note that

$$(U^{-1}w)_s = P(\lambda_s; D)U^{-1}w = P(\lambda_s; D)((e^D - I)U^{-1}b + g(1; D)U^{-1}c).$$

We have $P(\lambda_s; D)(e^D - I) = P(\lambda_s; D)(e^{\lambda_s} - 1)I$. In particular, when $\lambda_s = 0$, $P(\lambda_s; D)(e^D - I) = 0$. $P(\lambda_s; D)g(1; D) = \frac{e^{\lambda_s} - 1}{\lambda_s}P(\lambda_s; D)$ if $\lambda_s \neq 0$, and otherwise $P(\lambda_s; D)g(1; D) = P(\lambda_s; D)$. The formula is then proved. We next apply the arguments in remark 1 of theorem 1 to obtain the desired conclusions. \square

Furthermore, if A has a real eigenvalue λ , then one can recover λ from e^λ . However, this is not true for complex eigenvalues, since e^z is not an injective mapping on the complex domain \mathbb{C} .

3. Numerical algorithms

In this section, we propose various algorithms to recover the eigenvalues of A from data. Recall that proposition 1 in fact provides us with a Prony-type algorithm to reconstruct the annihilating polynomial $q_{S_\Omega, b}^A$. The original Prony method was proposed centuries ago to recover a vector with an s -sparse Fourier transform from $2s$ of its consecutive components. In recent years, the Prony method has been widely applied to different inverse problems including the approximation of Green functions in fluid dynamics [11], the inverse scattering problem [24], the parameter estimation in signal processing [13]. The renaissance of Prony method originates from its stabilized variants, such as the ESPRIT method [40], the matrix pencil method [27] or the approximate Prony method [37, 39, 45]. Error estimates for the performance of Prony-type methods with noisy measurements can be found in [9, 10, 22, 38, 47]. The robustness depends on the minimal gap between parameters.

To apply Prony-type methods, we first construct Hankel-like matrix. Let $\Omega \subset [d]$ and given the trajectory data $\{S_\Omega A^t b, t = 0, \dots, M-1\}$, the rectangular Hankel-like matrices for some

positive integer $L < M$ is given by

$$H_{\Omega, M-L, L+1} = \begin{bmatrix} | & | & & | \\ h_0 & h_1 & \dots & h_L \\ | & | & & | \end{bmatrix}, \quad (33)$$

$$H_{\Omega, M-L, L}(t) = H_{\Omega, M-L, L+1}(:, t+1:t+L), \quad t = 0, 1,$$

where the $(M-L)|\Omega| \times 1$ column vector is

$$h_l = [S_{\Omega}A^l b, S_{\Omega}A^{l+1} b, \dots, S_{\Omega}A^{M-L+l-1} b]^T, \quad (34)$$

for $l = 0, 1, \dots, L$. Denote by $r = r_{S_{\Omega}, b}^A$, the degree of minimal (S_{Ω}, A, b) annihilator $q_{S_{\Omega}, b}^A$. As long as $M \geq L+r$ and $L \geq r$, we have that $\text{rank}(H_{M-L, L}) = r$ and we can recover all Ω -recoverable eigenvalues. A practical choice of L is $L \lesssim M/2$ (and $L \leq N/2$), which is the maximum value for L . In particular, when $L = r$, the Hankel-like matrix is full column rank.

Determining the number of Ω -recoverable eigenvalues. In applications, the degree r of the annihilator $q_{S_{\Omega}, b}^A$ is sometimes not known as *a priori*. The information of r is crucial in applying the Prony-typed method proposed in proposition 1.

Algorithm 1. Determining the number of Ω -recoverable eigenvalues.

Input: $M \in \mathbb{N}$ s.t. $M > 3d$, observed samples $\{S_{\Omega}A^t b, t = 0, \dots, M-1\}$.

Step 1. Construct Hankel like matrix. Construct $H_{\Omega, M-d, d+1}$ from (33) and compute its singular values in descending order by taking SVD as $H_{\Omega, M-d, d+1} = U_{\Omega, M-d} D_{\Omega, M-d, d+1} W_{\Omega, d+1}^*$.

Step 2. Determine the numerical rank. We have different methods to determine the numerical rank including: (i) count the maximum number of singular values r_1 such that $\sigma_i \in \mathbb{R}$, $i = 1, 2, \dots, r$ are greater than a given positive threshold ϵ . (ii) Compute the quotient of singular values as $q_r = \frac{\sigma_r}{\sigma_{r+1}}$ and find the maximum quotient index $r_2 = \max_r q_r$. (iii) Reorder the quotient q_r in descending order as \hat{q}_r , find the maximum gap $\hat{q}_r - \hat{q}_{r+1}$, and determine the original order of \hat{q}_r in quotients $\{q_r\}$ as r_3 .

Output: the number of recoverable eigenvalues r_1, r_2, r_3 produced by three criterions.

We can estimate r by calculating the numerical rank of the Hankel matrix $H_{\Omega, M-L, L+1}$, given sufficient large L and M . For example, $L = d$ and $M \geq 2d$. We then compute the singular value decomposition (SVD) of $H_{\Omega, M-L, L+1}$ and find r as the number of singular values being larger than a threshold ϵ , which could be a fixed constant or the largest spectral gap among singular values, see algorithm 1 for more details.

In the following subsections, we will further investigate the data structure of the Hankel-like matrix $H_{\Omega, M-L, L+1}$ and present various algorithms based on the ideas from classical Prony, matrix pencil method, and ESPRIT estimation method. The latter two classical methods are well known for their better numerical stability than the original Prony method.

3.1. The Prony algorithm

The Prony-type method contained in the proof of proposition 1 is summarized as follows.

Algorithm 2. Prony LS method.

Input: $M \in \mathbb{N}$, s.t. $M > 3d$, observed samples $\{S_{\Omega}A^t b, t = 0, \dots, M-1\}$ and the number of recoverable eigenvalues r .

Step 1. Construct and solve Hankel-like matrix equation. Construct $H_{\Omega,M-r,r+1}$ from (33) and compute its SVD as $H_{\Omega,M-r,r+1} = U_{\Omega,M-r} D_{\Omega,M-r,r+1} W_{\Omega,r+1}^*$. Construct the linear equation in (21): $H_{\Omega,M-r,r}(0)\vec{q} = -h_{\Omega,r}$ from (33) and (34) and compute its least square solution and total least square solution $\vec{q} = (q_0, q_1, \dots, q_{r-1})^\top$ as

$$\vec{q}^{\text{LS}} = -H_{\Omega,M-r,r}(0)^+ h_{\Omega,r+1}.$$

Step 2. Find the roots of the polynomial. Compute all zeros $z_j^{\text{LS}} \in \mathbb{C}$, $j = 1, 2, \dots, r$ of $q(z) = \sum_{k=0}^{r-1} q_k z^k + z^r$ with \vec{q}^{LS} .

Output: $\{z_j^{\text{LS}} \in \mathbb{C}, j = 1, 2, \dots, r\}$.

3.2. Generalized matrix pencil method

In this section, we show that one can generalize the idea of matrix pencil to recover the eigenvalues of A . We first present a decomposition of Hankel matrices $H_{\Omega,M-L,L}(t)$, $t = 0, 1$ for the case when Ω has a single point.

Lemma 2. Let $\Omega = \{i\} \subset [d]$. Denote by $r = r_{S_\Omega,b}^A$ and $r_s = r_{S_\Omega,b_s}^A$ for $s = 1, \dots, n$. Then the rectangular Hankel matrices can be factorized in the following form for $t = 0, 1$:

$$H_{M-L,L}(t) = V_{r,M-L}^T \Lambda \hat{J}^t V_{r,L}, \quad (35)$$

$$V_{r,L} := \begin{bmatrix} V_{r_1,L} \\ V_{r_2,L} \\ \vdots \\ V_{r_n,L} \end{bmatrix}, \quad V_{r_s,L} = \begin{bmatrix} C_0^0 & C_1^0 \lambda_s & \dots & C_{r_s-1}^0 \lambda_s^{r_s-1} & \dots & C_{L-1}^0 \lambda_s^{L-1} \\ 0 & C_1^1 & \dots & C_{r_s-1}^1 \lambda_s^{r_s-2} & \dots & C_{L-1}^1 \lambda_s^{L-2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_{r_s-1}^{r_s-1} \lambda_s^0 & \dots & C_{L-1}^{r_s-1} \lambda_s^{L-r_s} \end{bmatrix} \in \mathbb{C}^{r_s \times L},$$

and an r -by- r matrix Λ :

$$\Lambda := \begin{bmatrix} \Lambda_1 & 0 & \dots & 0 \\ 0 & \Lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_n \end{bmatrix}, \quad \Lambda_s = \begin{bmatrix} \langle (U^{-1}b)_s, (U^* e_i)_s \rangle & \dots & \langle N_s^{r_s-1}(U^{-1}b)_s, (U^* e_i)_s \rangle \\ \langle N_s(U^{-1}b)_s, (U^* e_i)_s \rangle & \dots & 0 \\ \vdots & \vdots & \vdots \\ \langle N_s^{r_s-1}(U^{-1}b)_s, (U^* e_i)_s \rangle & \dots & 0 \end{bmatrix} \in \mathbb{C}^{r_s \times r_s},$$

and a Jordan matrix \hat{J} :

$$\hat{J} := \begin{bmatrix} \lambda_1 + \hat{N}_1 & 0 & \dots & 0 \\ 0 & \lambda_2 + \hat{N}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n + \hat{N}_n \end{bmatrix}, \quad \hat{N}_s = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \in \mathbb{C}^{r_s \times r_s}.$$

Proof. We prove this lemma by the matrices in (35) are entrywise identical. The (m, l) th entry of the Hankel matrix $H_{\Omega, M-L, L}(0)$ is given by

$$\begin{aligned} \langle A^{m+l-2}b, e_i \rangle &= \langle J^{m+l-2}(U^{-1}b), U^*e_i \rangle = \sum_{s=1}^n \langle J_s^{m+l-2}(U^{-1}b)_s, (U^*e_i)_s \rangle \\ &= \sum_{s=1}^n \sum_{k_s=0}^{r_s-1} \binom{m+l-2}{k_s} \lambda_s^{m+l-2-k_s} \\ &\quad \times \langle N_s^{k_s}(U^{-1}b)_s, (U^*e_i)_s \rangle. \end{aligned}$$

Then by using the identity

$$\binom{m+l-2}{k_s} = \sum_{i=0}^{k_s} \binom{m-1}{i} \binom{l-1}{k_s-i},$$

and comparing the coefficients for $\langle N_s^i(U^{-1}b)_s, (U^*e_i)_s \rangle$ for $i = 0, \dots, r_s - 1$, one can show that it is the same with the (m, l) th entry of the matrix $V_{r, M-L}^T \Lambda V_{r, L}$. Similarly, we can prove for the Hankel matrix $H_{\Omega, M-L, L}(1)$. \square

Recall that the superscripts ‘*’ and ‘+’ will denote the conjugate transpose and the pseudoinverse. The following lemma provides a theoretical foundation for the generalized matrix pencil method.

Proposition 2. Let $\Omega = \{i\} \subset [d]$. Denote by $r = r_{S_\Omega, b}^A$ and $r_s = r_{S_\Omega, b_s}^A$ for $s = 1, \dots, n$. Without loss of generality, assume that $r_s \geq 1$. Let M, L be two positive integers such that $r \leq L \leq M - r$. The solutions to the generalized singular eigenvalue problem:

$$(zH_{M-L, L}(0) - H_{M-L, L}(1))v = 0 \quad (36)$$

subject to $b \in \text{Col}(H_{\Omega, M-L, L}^*(0))$, denoting the column space of $H_{\Omega, M-L, L}^*(0)$, are

$$z_s = \lambda_s,$$

$$v = \text{the } r_1 + \dots + r_s \text{th column of } V_{r, L}^+,$$

for $s = 1, \dots, n$.

Proof. Using the factorization (35), we re-write the equation (36) as

$$V_{r, M-L}^T \Lambda (z - \hat{J}) V_{r, L} v = 0.$$

Since the matrix $V_{r, L}$ has linearly independent rows, every x in $\text{Col}(H_{\Omega, M-L, L}^*(0))$ can be represented as $v = V_{r, L}^+ c$ for some vector $c \in \mathbb{C}^r$ due to the property of pseudoinverse. We also note that the matrix $V_{r, M-L}^T$ has linearly independent columns, and the matrix Λ is invertible (by proposition 1). Therefore, it suffices to solve

$$(z - \hat{J})c = 0.$$

We then know the possible values for z such that the linear equations have nonzero solutions are $\lambda_1, \dots, \lambda_n$. The conclusion follows by solving the corresponding linear system of equations. \square

The general case. Let $\Omega = \{i_1, \dots, i_k\} \subset [d]$, and assume that $r_s = r_{S_{\Omega,b}^A} \geq 1$ for $s = 1, \dots, n$. By appropriate permutations of rows, the Hankel-like matrix $H_{\Omega,M-L,L+1}$ defined in (33) can be transformed as

$$\tilde{H}_{\Omega,M-L,L+1} := \begin{bmatrix} H_{\{i_1\},M-L,L+1} \\ \vdots \\ H_{\{i_k\},M-L,L+1} \end{bmatrix}. \quad (37)$$

Then each Hankel matrix $H_{\{i_j\},M-L,L+1}$ can be still factorized in the form of (35)

$$H_{\{i_j\},M-L,L}(t) = V_{r,M-L}^T \Lambda_{\{i_j\}} \hat{J}^t V_{r,L}, \quad t = 0, 1. \quad (38)$$

In particular, if $r_{S_{\{i_j\},b}^A} = 0$, then the corresponding $\Lambda_{\{i_j\},s}$ in (2) will be a $r_s \times r_s$ zero matrix.

The solutions to the generalized singular eigenvalue problem:

$$(zH_{\Omega,M-L,L}(0) - H_{\Omega,M-L,L}(1))v = 0, \quad (39)$$

subject to $x \in \text{Col}(H_{\Omega,M-L,L}^*(0))$, denoting the column space of $H_{\Omega,M-L,L}^*(0)$ are the same with the problem

$$(z\tilde{H}_{\Omega,M-L,L}(0) - \tilde{H}_{\Omega,M-L,L}(1))\tilde{v} = 0, \quad (40)$$

subject to $\tilde{v} \in \text{Col}(\tilde{H}_{\Omega,M-L,L}^*(0))$, in the sense that \tilde{v} is a permutation of v . Due to the factorization (35) and that $r_s = \max_{j=1, \dots, k} r_{S_{\{i_k\},b}^A}$, we have that $\text{Col}(\tilde{H}_{\Omega,M-L,L}^*(0)) = \text{Range}(V_{r,L}^+)$. Therefore the solution to (40) is equivalent to

$$\Lambda_{\{i_j\}}(z - \hat{J})\tilde{c} = 0, \quad j = 1, \dots, k. \quad (41)$$

Since for each $s = 1, \dots, n$, there is at least one $\Lambda_{\{i_j\},s}$ in (2) is invertible, therefore the values of z to solve (41) with nonzero \tilde{c} are $\lambda_1, \dots, \lambda_n$. We therefore obtain the following conclusion:

Theorem 5. Let $\Omega \subset [d]$, and denote $r := r_{S_{\Omega,b}^A}$. Let M, L be two positive integers such that $r \leq L \leq M - r$. The $L \times L$ matrix $H_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(1)$ has the same eigenvalues with roots of $q_{S_{\Omega,b}^A}$ and $L - r$ zeros as eigenvalues.

Proof. Left multiplying (39) by $H_{\Omega,M-L,L}^*(0)$, we have

$$H_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(1)x = zH_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(0)x. \quad (42)$$

By property of the pseudoinverse, $H_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(0)$ is the orthogonal projection onto $\text{Col}(H_{\Omega,M-L,L}^*(0))$. Since $x \in \text{Col}(H_{\Omega,M-L,L}^*(0))$, it is easy to see that $\lambda_1, \dots, \lambda_n$ are n eigenvalues of $H_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(1)$. Since the rank of $H_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(1)$ is $r \leq L$, $H_{\Omega,M-L,L}^*(0)H_{\Omega,M-L,L}(1)$ has $L - r_{S_{\Omega,b}^A}$ zero eigenvalues. \square

One advantage of the matrix pencil method is the fact that there is no need to compute the coefficients of the minimal annihilating polynomial $q_{S_{\Omega,b}^A}$. In this way, we need only solve a

standard eigenvalue problem of a square matrix $H_{\Omega,M-L,L}^+(0)H_{\Omega,M-L,L}(1)$. In order to compute $H_{\Omega,M-L,L}^+(0)H_{\Omega,M-L,L}(1)$. Inspired by the idea of algorithm 5 in [28], we can employ the SVD based matrix pencil method for Hankel-like matrices.

Proposition 3. *In addition to the conditions of theorem 5, given the SVD of the Hankel-like matrix,*

$$H_{\Omega,M-L,L+1} = U_{\Omega,M-L}\Sigma_{\Omega,M-L,L+1}W_{\Omega,L+1}^*,$$

then

$$H_{\Omega,M-L,L}^+(0)H_{\Omega,M-L,L}(1) = (W_{\Omega,L+1}^*(1:r, 1:L))^+ (W_{\Omega,L+1}^*(1:r, 2:L+1)).$$

We now summarize the generalized matrix pencil method as below.

Algorithm 3. Matrix pencil method

Input: $M \in \mathbb{N}$ s.t. $M > 3d$, observed samples $\{S_{\Omega}A^t b, t = 0, \dots, M-1\}$ and the number of recoverable eigenvalues r .

Step 1. Construct and solve Hankel-like matrix equation. Construct the matrix equation in as $H_{\Omega,M-L,L}(0)C = H_{\Omega,M-L,L}(1)$ from equations (33) and (34) and compute SVD of augmented matrix $[H_{\Omega,M-L,L}(0) \ H_{\Omega,M-L,L}(1)] = U'_{\Omega,M-L}D'_{\Omega,M-L,2L}W_{\Omega,2L}^{*}$ and compute its least square solution, total least square solution and SVD-based solution by

$$\begin{aligned} C^{\text{LS}} &= (H_{\Omega,M-L,L}(0))^+ H_{\Omega,M-L,L}(1). \\ C^{\text{SVD}} &= (W_{\Omega,L+1}^*(1:r, 1:L))^+ W_{\Omega,L+1}^*(1:r, 2:L+1). \end{aligned}$$

Step 2. Find the eigenvalues of companion matrix. Compute all eigenvalues $\lambda_j^{\text{LS}}, \lambda_j^{\text{SVD}} \in \mathbb{C}, j = 1, 2, \dots, L$ of the companion matrix $C^{\text{LS}}, C^{\text{SVD}}$, respectively.

Step 3. Remove redundant zeros. Remove the eigenvalues with sufficiently small norm, i.e., $\|\lambda_j\| \leq \eta$ for a small positive number $\eta \in \mathbb{R}^+$.

Output: $\{z_j^{\text{LS}} \in \mathbb{C}, j = 1, 2, \dots, L | \|\lambda_j^{\text{LS}}\| > \eta\}, \{z_j^{\text{SVD}} \in \mathbb{C}, j = 1, 2, \dots, L | \|\lambda_j^{\text{SVD}}\| > \eta\}$.

3.3. Generalized ESPRIT method

The original ESPRIT method relies on a particular property of Vandermonde matrices known as rotational invariance [40]. By the factorization (35), and the permutation argument in (37), we have seen that the Hankel-like data matrix $H_{\Omega,M-L,L+1}$ is rank-deficient and that its range space, spanned by columns of $V_{\Omega,r,M-L}^T$, satisfies a generalized rotation invariance property:

$$V_{\Omega,r,M-L}^T(2:M-L,:) = V_{\Omega,r,M-L}^T(1:M-L-1,:)\hat{J}^T, \quad (43)$$

where the matrix $V_{\Omega,r,M-L}$ and the Jordan matrix \hat{J} is defined in (35). Hence, we can generalize the ESPRIT algorithm based on SVD for estimating the eigenvalues of \hat{J} in our setting. We summarize this method below:

Algorithm 4. ESPRIT method

Input: $M \in \mathbb{N}$ s.t. $M > 3d$, observed samples $\{S_{\Omega}A^t b, t = 0, \dots, M-1\}$ and the number of recoverable eigenvalues r .

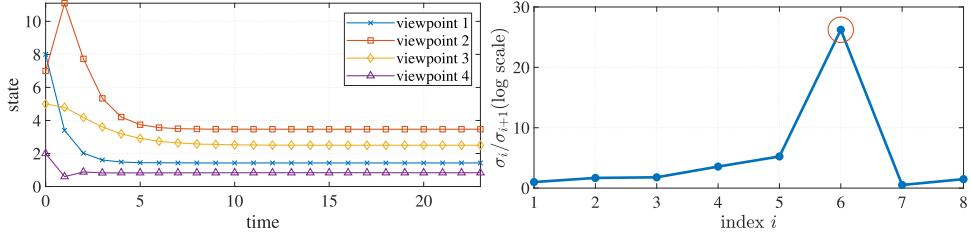


Figure 1. Example 1. (Left) State-time plot of observations of the trajectory with $\Omega = \{1, 2, 4, 7\}$. (Right) Spectrum quotient of Hankel-type matrix for the choice of $\Omega = \{1, 2, 4, 7\}$.

Step 1. Construct and solve matrix equation. Construct the generalized rotation invariance matrix equation as in (43) and compute its solution by

$$\hat{J} = (U_{\Omega, M-L}(1 : M-L-1, 1 : r))^+ U_{\Omega, M-L}(2 : M-L, 1 : r).$$

Step 2. Find the eigenvalues of matrix. Compute all eigenvalues $\lambda_j^{\text{ES}} \in \mathbb{C}$, $j = 1, 2, \dots, r$ of matrix \hat{J} .

Output: $\{\lambda_j^{\text{ES}} \in \mathbb{C}, j = 1, 2, \dots, r\}$.

4. Empirical evaluations

In this section, we examine and compare the performance of algorithms 1–4 on estimating spectrum of various affine systems (see section 3). Our focus is the noise-free data and consider the system matrix A whose operator norm is no greater than 1. Denote by r the number of recoverable eigenvalues, and we let $\Lambda = \{\lambda_1, \dots, \lambda_r\}$ be Ω -recovered eigenvalue set, and $\hat{\Lambda} = \{\hat{\lambda}_1, \dots, \hat{\lambda}_r\}$ be the eigenvalue set obtained by our numerical algorithms. The performance of the algorithms is measured by the root mean squared error (RMSE) and the infinity norm error (INE) defined as.

$$\text{RMSE}(\Omega, M, \sigma) = \sqrt{\frac{1}{r} \sum_{i=1}^r (\lambda_i - \hat{\lambda}_i)^2}, \quad (44)$$

$$\text{RINF}(\Omega, M, \sigma) = \max_{i=1, \dots, r} |\lambda_i - \hat{\lambda}_i|. \quad (45)$$

In each example, we obtain the observational trajectory data of the form $\{S_\Omega x_t : t = 0, 1, \dots, M-1\}$, where $\Omega \subset [d]$. In numerical experiments, we get an estimation \hat{r} of the number of recoverable eigenvalues using algorithm 1 and then use the Hausdorff distance to match the recoverable eigenvalues with the exact ones as

$$\lambda_i = \arg \min_{\lambda \in \Lambda} (\lambda - \hat{\lambda}_i), \quad \text{for } \hat{\lambda}_i \in \hat{\Lambda}. \quad (46)$$

The results and details are listed below.

Table 1. Example 1. Numerical rank and errors for various algorithms and choices of Ω .

Ω	M	r	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
				RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
{1}	24	1	1	3.9×10^{-17}	1.1×10^{-16}	3.9×10^{-17}	1.1×10^{-16}	2.0×10^{-17}	5.6×10^{-17}	2.0×10^{-17}	5.6×10^{-17}
{1, 4}	24	4	4	4.3×10^{-8}	8.7×10^{-8}	1.9×10^{-8}	3.8×10^{-8}	4.4×10^{-8}	8.7×10^{-8}	1.7×10^{-8}	3.4×10^{-8}
{1, 4, 7}	24	5	5	6.0×10^{-8}	1.2×10^{-7}	4.0×10^{-8}	8.0×10^{-8}	3.3×10^{-8}	6.5×10^{-8}	4.8×10^{-4}	9.7×10^{-4}
{1, 2, 4, 7}	24	6	6	5.5×10^{-8}	9.9×10^{-8}	1.5×10^{-8}	2.9×10^{-8}	5.7×10^{-8}	9.5×10^{-8}	4.8×10^{-4}	9.7×10^{-4}

Table 2. Example 1. Errors for various algorithms and choices of M .

Ω	M	r	\hat{r}	Prony		Matrix pencil LS		Matrix pencil SVD		ESPRIT	
				RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
$\{1, 2, 4, 7\}$	24	6	6	5.5×10^{-8}	9.9×10^{-8}	1.5×10^{-8}	2.9×10^{-8}	5.7×10^{-8}	9.5×10^{-8}	4.8×10^{-4}	9.7×10^{-4}
$\{1, 2, 4, 7\}$	32	6	6	8.7×10^{-8}	1.5×10^{-7}	9.1×10^{-8}	1.7×10^{-7}	6.2×10^{-8}	1.2×10^{-7}	4.3×10^{-6}	8.7×10^{-6}
$\{1, 2, 4, 7\}$	40	6	6	1.1×10^{-7}	2.2×10^{-7}	7.6×10^{-8}	1.5×10^{-7}	1.3×10^{-7}	2.5×10^{-7}	8.7×10^{-8}	1.7×10^{-7}
$\{1, 2, 4, 7\}$	48	6	6	1.1×10^{-7}	2.0×10^{-7}	8.3×10^{-8}	1.6×10^{-7}	1.1×10^{-7}	2.1×10^{-7}	8.3×10^{-8}	1.6×10^{-7}

Table 3. Example 2. Numerical rank and errors of various algorithms and choices of Ω .

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
{1}	60	10	2.8×10^{-2}	8.1×10^{-2}	4.0×10^{-5}	1.3×10^{-4}	2.8×10^{-2}	8.1×10^{-2}	2.7×10^{-2}	7.8×10^{-2}
{1, 2}	60	19	7.6×10^{-3}	3.3×10^{-2}	4.2×10^{-5}	1.8×10^{-4}	7.6×10^{-3}	3.3×10^{-2}	1.5×10^{-1}	3.1×10^{-1}
{1, 2, 3}	60	20	7.3×10^{-7}	2.0×10^{-6}	7.1×10^{-6}	3.1×10^{-5}	1.2×10^{-6}	5.1×10^{-6}	8.9×10^{-2}	1.8×10^{-1}
{1, 2, 3, 4}	60	20	2.0×10^{-7}	7.7×10^{-7}	6.3×10^{-7}	2.7×10^{-6}	1.1×10^{-7}	3.9×10^{-7}	6.7×10^{-2}	1.3×10^{-1}

Table 4. Example 2. Errors of various algorithms and choices of M .

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
$\{1, 2, 3, 4\}$	60	20	2.0×10^{-7}	7.7×10^{-7}	6.3×10^{-7}	2.7×10^{-6}	1.1×10^{-7}	3.9×10^{-7}	6.7×10^{-2}	1.3×10^{-1}
$\{1, 2, 3, 4\}$	80	20	2.7×10^{-7}	1.1×10^{-6}	1.3×10^{-7}	5.8×10^{-7}	4.9×10^{-7}	2.2×10^{-6}	6.3×10^{-2}	1.3×10^{-1}
$\{1, 2, 3, 4\}$	100	20	1.3×10^{-6}	5.7×10^{-6}	2.5×10^{-7}	1.0×10^{-6}	6.1×10^{-7}	1.7×10^{-6}	5.4×10^{-2}	1.2×10^{-1}
$\{1, 2, 3, 4\}$	120	20	8.5×10^{-7}	3.7×10^{-6}	2.9×10^{-6}	1.3×10^{-5}	1.7×10^{-6}	7.6×10^{-6}	4.6×10^{-2}	8.8×10^{-2}

Table 5. Example 3. Numerical rank and errors of various algorithms and choices of Ω .

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
{1}	90	6	6.5×10^{-3}	2.7×10^{-2}	1.8×10^{-3}	7.2×10^{-3}	6.0×10^{-3}	2.4×10^{-2}	5.7×10^{-3}	2.2×10^{-2}
{1, 2}	90	7	4.0×10^{-3}	1.9×10^{-2}	2.2×10^{-3}	1.2×10^{-2}	4.2×10^{-3}	2.0×10^{-2}	1.5×10^{-2}	5.4×10^{-2}
{1, 2, 3}	90	13	3.7×10^{-6}	1.4×10^{-5}	1.5×10^{-5}	7.1×10^{-5}	1.2×10^{-5}	4.7×10^{-5}	5.2×10^{-2}	1.3×10^{-1}
{1, 2, 3, 4}	90	12	3.6×10^{-3}	1.2×10^{-2}	1.3×10^{-5}	6.2×10^{-5}	4.1×10^{-3}	1.7×10^{-2}	1.6×10^{-2}	5.4×10^{-2}
{1, 2, 3, 4, 5}	90	13	1.8×10^{-7}	8.4×10^{-7}	2.3×10^{-6}	1.2×10^{-5}	1.7×10^{-6}	8.8×10^{-6}	1.8×10^{-2}	6.3×10^{-2}

Table 6. Example 3. Errors of various algorithms and choices of M .

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
$\{1, 2, 3, 4, 5\}$	90	13	1.8×10^{-7}	8.4×10^{-7}	2.3×10^{-6}	1.2×10^{-5}	1.7×10^{-6}	8.8×10^{-6}	1.8×10^{-2}	6.3×10^{-2}
$\{1, 2, 3, 4, 5\}$	120	13	1.2×10^{-6}	5.9×10^{-6}	3.0×10^{-6}	1.4×10^{-5}	2.4×10^{-6}	1.2×10^{-5}	1.4×10^{-2}	4.8×10^{-2}
$\{1, 2, 3, 4, 5\}$	150	13	2.2×10^{-6}	1.1×10^{-5}	1.0×10^{-6}	5.4×10^{-6}	3.1×10^{-6}	1.5×10^{-5}	1.1×10^{-2}	3.0×10^{-2}
$\{1, 2, 3, 4, 5\}$	180	13	5.1×10^{-7}	2.1×10^{-6}	2.8×10^{-6}	1.4×10^{-5}	3.1×10^{-6}	1.6×10^{-5}	8.3×10^{-3}	2.4×10^{-2}

Example 1. Discrete affine system. Recall that a discrete state-time affine system is given by

$$x_{t+1} = Ax_t + c.$$

We consider a system of dimension 8. Denote by $J = \text{diag}(0.3I_3 + \hat{N}_3, 0.5I_2 + \hat{N}_2, 0.6, -0.2I_2)$, and $U = \text{diag}(I_3, \text{toeplitz}([1, 0, 0], [1, 1, 1]), \text{hankel}([1, 2], [2, 1]))$, where I_s denotes the $s \times s$ identity matrix, and \hat{N}_s denotes $s \times s$ nilpotent matrix with one cyclic block as in lemma 2. We have that $A = UJU^{-1}$ with the initial condition $x_0 = [8, 7, \dots, 1]^T$ and $c = [1, 1, \dots, 1]^T$. To illustrate the reconstruction flows, we first depict the observations of the discrete affine system for the choice of $\Omega = \{1, 2, 4, 7\}$ in figure 1 (left). Note that by introducing $y_t = x_t + (A - I_8)^{-1}c$, the discrete affine system is reduced to the linear dynamical system. We can then apply algorithm 1 to construct Hankel-type matrix $H_{\Omega, M-L, L}$ and compute its quotients of singular values σ_i/σ_{i+1} for $i = 1, 2, \dots, 8$ as shown in figure 1 (right). In this case, six eigenvalues are Ω -recoverable, matching the analytical results in theorem 1. To further compare the performance of various algorithms and investigate the impact of Ω, M , we conduct the reconstruction with various choices of Ω and M as shown below in tables 1 and 2.

Dynamical processes on graphs. Graph learning arises in a wide range of applications. We consider a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ in which $\mathcal{V} = \{v_1, \dots, v_d\}$ is set of d vertices and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a set of edges. The weighted adjacent matrix \mathcal{W} is defined as

$$\mathcal{W}(i, j) \triangleq \begin{cases} \alpha_{ij} & \text{if the directed pair } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}; \alpha_{ij} \in \mathbb{R}_+; \forall v_i, v_j \in \mathcal{V}. \quad (47)$$

The degree $\deg(v_i)$ of a vertex $v_i \in \mathcal{V}$ is defined as $\deg(v_i) = \sum_{j=1}^n \mathcal{W}(i, j)$. In the following, we introduce important operators associated with the graph \mathcal{G} .

Definition 6. The normalized diffusion operator of a graph \mathcal{G} with the weighted adjacent matrix $\mathcal{W} \in \mathbb{R}^{d \times d}$ is defined by $\mathcal{A} \triangleq (\mathcal{D}^{-1})^{\frac{1}{2}} \mathcal{W} (\mathcal{D}^{-1})^{\frac{1}{2}}$, where $\mathcal{D} := \text{diag}(\deg(v_i))_{v_i \in \mathcal{V}}$ and \mathcal{D}^{-1} denote its psedoinverse. The normalized graph Laplacian operator is $\mathcal{L} = I - \mathcal{A}$. The random walk transition matrix \mathcal{P} is defined by $\mathcal{D}^{-1} \mathcal{W}$.

Remark 2. A weighted graph is called simple graph if $\mathcal{W}(i, j) \in \{0, 1\}$ and $\mathcal{W}(i, i) = 0$. In spectral graph theory, other types of Laplacian matrix are also considered. For example, the combinatorial Laplacian matrix $\{\hat{\mathcal{L}}\} = \mathcal{D} - \mathcal{A}$. The signless Laplacian matrix $|L| = \mathcal{D} + \mathcal{A}$. It is possible to use any type of Laplacian matrix in the heat kernel for the diffusion process. In our numerical sections, we shall restrict our attention to the simple weighted graph and the normalized graph Laplacian. Our theorems and algorithms can be also applied to other types of Laplacian matrices too.

Note that if the vertex v_i in a graph is isolated, then the degree matrix $\mathcal{D}(i, i) = 0$. In this case, we use psuedo-inverse of \mathcal{D} to calculate the transition matrix \mathcal{P} and set $\mathcal{P}(i, i) = 1$.

(a) Random walk over graphs. A random walk on graph is a dynamical process comprised of a series of random steps by moving to an adjacent vertex at each step: if $v(t)$ represents the vertex of the random walk at the timestep t then we moves to the next one $v(t + 1)$ by picking

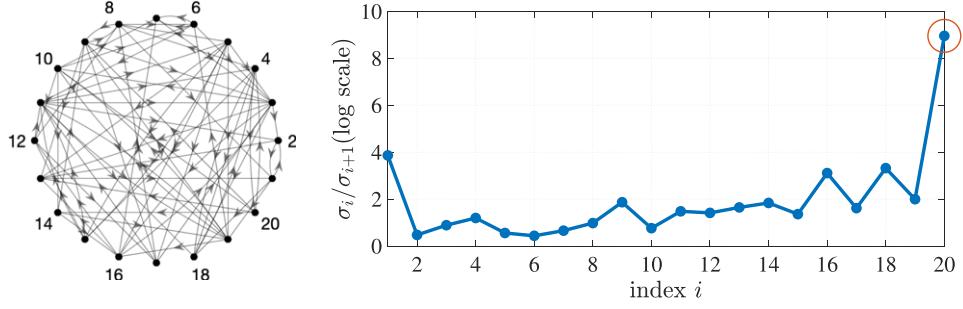


Figure 2. Example 2. (Left) An illustration of vertices and edges in the directed unweighted graph. (Right) Spectrum quotient of Hankel-type matrix for the choice of $\Omega = \{1, 2, \dots, 7\}$. In this case, 20 eigenvalues are Ω -recoverable.

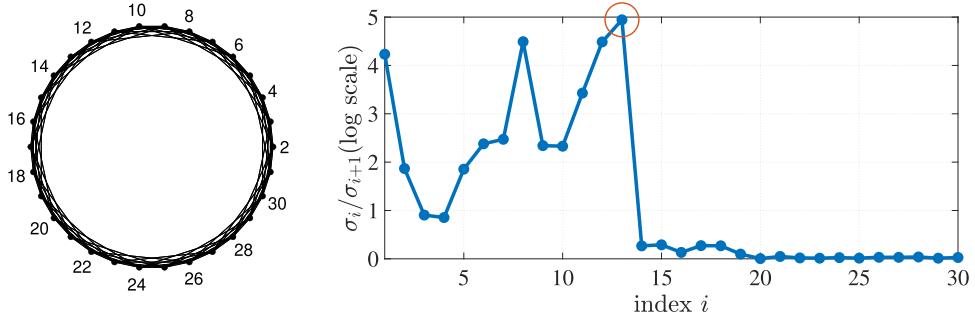


Figure 3. Example 3. (Left) An illustration of vertices and edges in the ring-shaped graph. (Right) Spectrum quotient of Hankel-type matrix for the choice of $\Omega = \{1, 2, \dots, 5\}$. In this case, 13 eigenvalues are Ω -recoverable.

one of its neighbors with probability,

$$\mathbb{P}(v(t+1)|v(t)) = \begin{cases} \frac{1}{\deg(v(t))}, & \text{if } (v(t), v(t+1)) \in \mathcal{E}, \\ 0, & \text{otherwise,} \end{cases} \quad (48)$$

where $\deg(v(t))$ denotes the number of edges starting from $v(t)$ in digraph \mathcal{G} . Let x_t denote the probability distribution at time t

$$x_t(i) = \mathbb{P}(v(t) = v_i).$$

By rule of random walk, we have the following linear evolution system,

$$x_t = A^t x_0, \quad A = \mathcal{P}^T.$$

The eigenvalues of A reveals useful information about the underlying graph: the multiplicity of eigenvalue 1 is equal to the number of (strongly) connected components; the second largest eigenvalue λ_2 that describes the mixing rate of the random walks; the spectral gap $|\lambda_1 - \lambda_2|$ represents how well the graph is connected. The reader can refer to the book [17] for more connections between them (tables 3, 4, 5 and 6).

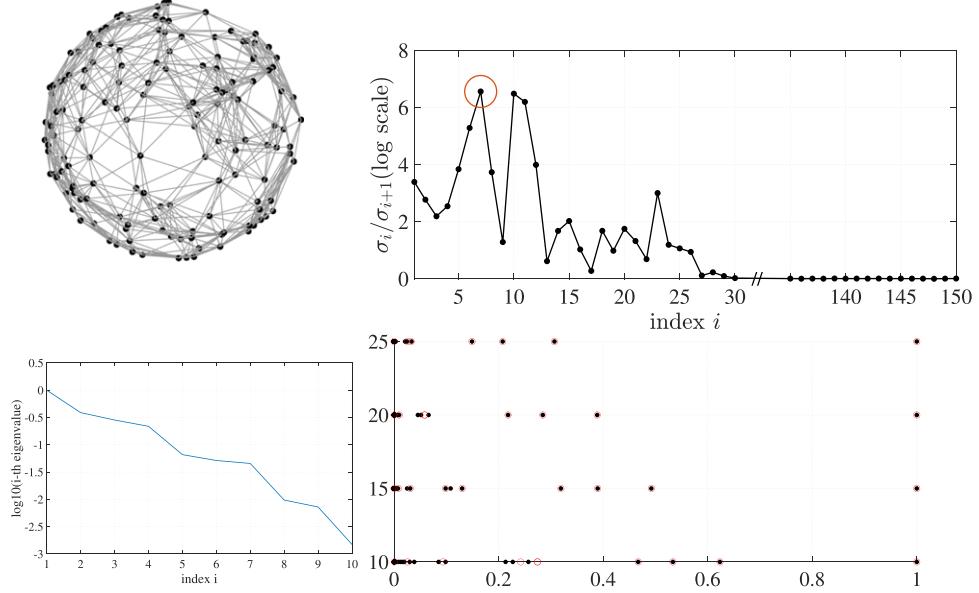


Figure 4. Example 4. (Top left) An illustration of vertices and edges in the sphere-shaped graph. (Top right) Spectrum quotient of Hankel-type matrix for the choice of $\Omega = \{1, 2, 3, 4\}$. In this case, seven eigenvalues are Ω -recoverable. (Bottom left) The eigenvalue of A when $\Delta t = 20$. (Bottom right) The ground truth spectrum (filled points) and reconstructed spectrum (hollow points) for $\Delta t = 10 : 5 : 25$ where we set the rank $\hat{r} = 10$ in all cases.

Example 2. A directed unweighted graph. We consider a simple directed unweighted graph of 20 nodes. Its weighted adjacent matrix \mathcal{W} (its nonzero entries are all 1s) is randomly generated with 80 edges and we then remove the self-loops. The initial state x_0 is a non-degenerate discrete probability distribution on $\{1, \dots, 20\}$. The matrix A in our example is an invertible diagonalizable matrix with 20 distinct eigenvalues and its eigen-matrix U is entrywise nonzero. We summarize the reconstruction results below. In particular, we recovered the multiplicity of 1 in this case, indicating that the graph has only one strongly connected component.

Example 3. Ring graph [35]. An undirected graph with 30 nodes uniformly distributed on a ring-shaped structure and each vertex has eight neighbors. The edge weights are all equal to 1. The matrix A in this example is a diagonalizable matrix with 13 distinct eigenvalues. Unlike example 1, in this case, there are no repeating roots in the annihilating polynomial and therefore we can not recover their geometric multiplicities by our algorithms. We summarize the results below.

(b) Heat diffusion process over graphs. We consider a non-homogeneous heat diffusion process over the graph \mathcal{G} is governed by a continuous affine system

$$x_{t+1}^{\text{cont}} = e^{-t\mathcal{L}} x_t^{\text{cont}} + c, \quad t \geq 0. \quad (49)$$

We observe the system at uniform time instances $0, \Delta t, 2\Delta t, \dots$. In this subsection, we explore the reconstruction of eigenvalues when the system matrix is relatively large (figures 2 and 3).

Table 7. Example 4. Numerical rank and errors of various algorithms and choices of Ω for $\Delta t = 20$.

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
{1, 2}	450	5	8.8×10^{-4}	8.3×10^{-3}	2.4×10^{-4}	2.8×10^{-3}	8.8×10^{-4}	8.3×10^{-3}	7.3×10^{-4}	6.3×10^{-3}
{1, 2, 3}	450	6	1.5×10^{-4}	1.8×10^{-3}	1.8×10^{-3}	1.6×10^{-2}	1.5×10^{-4}	1.8×10^{-3}	1.5×10^{-4}	1.8×10^{-3}
{1, 2, 3, 4}	450	7	7.8×10^{-4}	9.5×10^{-3}	6.8×10^{-4}	7.7×10^{-3}	7.8×10^{-4}	9.4×10^{-3}	7.6×10^{-4}	9.2×10^{-3}
{1, 2, 3, 4, 5}	450	7	8.1×10^{-4}	9.8×10^{-3}	9.4×10^{-4}	1.1×10^{-2}	8.1×10^{-4}	9.8×10^{-3}	7.8×10^{-4}	9.5×10^{-3}

Table 8. Example 4. Errors of various algorithms and choices of M .

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
$\{1, 2, 3, 4, 5\}$	450	7	8.1×10^{-4}	9.8×10^{-3}	9.4×10^{-4}	1.1×10^{-2}	8.1×10^{-4}	9.8×10^{-3}	7.8×10^{-4}	9.5×10^{-3}
$\{1, 2, 3, 4, 5\}$	600	7	8.1×10^{-4}	9.8×10^{-3}	9.4×10^{-4}	1.1×10^{-2}	8.1×10^{-4}	9.8×10^{-3}	7.8×10^{-4}	9.5×10^{-3}
$\{1, 2, 3, 4, 5\}$	750	7	8.1×10^{-4}	9.8×10^{-3}	9.4×10^{-4}	1.1×10^{-2}	8.1×10^{-4}	9.8×10^{-3}	7.8×10^{-4}	9.5×10^{-3}
$\{1, 2, 3, 4, 5\}$	900	7	8.1×10^{-4}	9.8×10^{-3}	9.4×10^{-4}	1.1×10^{-2}	8.1×10^{-4}	9.8×10^{-3}	7.8×10^{-4}	9.5×10^{-3}

Table 9. Example 5. Numerical rank and errors of various algorithms and choices of Ω .

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
{1}	30	3	7.7×10^{-13}	1.3×10^{-12}	2.7×10^{-13}	4.2×10^{-13}	9.7×10^{-13}	1.6×10^{-12}	8.8×10^{-14}	1.5×10^{-13}
{1, 2}	30	3	1.4×10^{-13}	1.4×10^{-13}	1.1×10^{-13}	1.1×10^{-13}	7.7×10^{-14}	8.1×10^{-14}	5.5×10^{-2}	6.6×10^{-2}
{1, 2, 3}	30	3	1.8×10^{-13}	2.0×10^{-13}	4.4×10^{-14}	5.3×10^{-14}	1.7×10^{-13}	2.1×10^{-13}	6.4×10^{-2}	7.7×10^{-2}

Table 10. Example 6. Root Mean square errors and infinity errors for various algorithms. In this table, \hat{r} is the estimator of algorithm 1 using (i) in step 2 with relative threshold $\epsilon = 10^{-3}$.

Ω	M	\hat{r}	Prony		Matrix pencil		Matrix pencil SVD		ESPRIT	
			RMSE	INE	RMSE	INE	RMSE	INE	RMSE	INE
{1}	316	10	2.2×10^{-1}	7.9×10^{-1}	2.3×10^{-1}	9.8×10^{-1}	1.6×10^{-2}	5.7×10^{-2}	2.0×10^{-2}	9.1×10^{-2}
{1, 3, 9}	316	12	2.6×10^{-1}	8.3×10^{-1}	3.5×10^{-1}	9.7×10^{-1}	1.9×10^{-2}	7.2×10^{-2}	3.6×10^{-2}	1.6×10^{-1}
{1, 2, 3, 9}	316	11	2.4×10^{-1}	7.7×10^{-1}	3.5×10^{-1}	9.7×10^{-1}	2.6×10^{-2}	1.3×10^{-1}	2.2×10^{-2}	8.8×10^{-2}
{1, 2, 3, 5, 9}	316	12	2.4×10^{-1}	7.3×10^{-1}	3.5×10^{-1}	9.7×10^{-1}	2.6×10^{-2}	9.0×10^{-2}	3.2×10^{-2}	1.4×10^{-1}

Example 4. Sphere graph [35]. An undirected graph with 150 nodes sampled on a hyper-sphere and each vertex is connected to its 10 nearest neighbors. In this example, $x_t^{\text{cont}}(0)$ is randomly generated from the uniform distribution and c is in the image of a random Gaussian vector under the map $e^{-t\mathcal{L}}$. We observe the system at $t_l = l\Delta t$ for $l = 0, \dots, M-1$ and $\Delta t = 20$. The matrix $A = e^{-20\mathcal{L}}$ is approximately low rank: only ten eigenvalues are greater than 10^{-3} (see the left bottom panel of figure 4). The largest gap happens between 7th eigenvalue 0.0455 and 8th eigenvalue 0.0097. We summarize the spectral plot of the numerical rank estimation (the top right panel of figure 4) and reconstruction results in the table 7 (various Ω) and table 8 (various M). It shows that our algorithms can recover significant eigenvalues very well. We also investigate various choice of Δt . When Δt is relatively small, all eigenvalues of A lie in $[0, 1]$ and form clusters. Two eigenvalues in the same cluster are very close to each other and can be identified with the same value ‘numerically’. Below, we show that our algorithms can recover the representative eigenvalues in each cluster (the right bottom panel of figure 4).

Approximate discrete linear dynamical system. In the last subsection, we consider the case where the time series data can be well-approximated by a discrete linear dynamical system governed by A . We use partial observations of the original time series data and compare the reconstruction results with the eigenvalues of A .

Example 5. Non-linear LIP model. In this example, we consider a three dimensional discrete homogeneous dynamical system that serves as an approximation to the LIP model of influenza virus inflection model:

$$\begin{cases} \dot{V} = rI - cV \\ \dot{H} = -\beta HV \\ \dot{I} = \beta HV - \delta I \end{cases} . \quad (50)$$

Similar to the paper [21], here we also use a rescaled model for convenience, with parameters $\beta = 10.8$, $r = 12$, $c = 3$, $\delta = 4$ and initial state $V(0) = 0.093/(4 \times 10^5)$, $H(0) = 1$, $I(0) = 0$. We use Matlab built-in function `ode45` to obtain the solution. The code can be found in XX (tables 9 and 10).

We consider the time series data when $t \in [1.8, 2.1]$ as the underlying noise-free observations (with $\Delta t = 0.01$), where the number of uninfected cells is rapidly decreasing due to the increasing number of infected cells and their released virus (see the top right panel of figure 5). We consider the least square approximation to obtain a discrete homogeneous linear system as

$$A = \begin{bmatrix} | & | & & | \\ x_0 & x_1 & \dots & x_{N-2} \\ | & | & & | \end{bmatrix}^\dagger \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_{N-1} \\ | & | & & | \end{bmatrix}, \quad (51)$$

where x_t for $t = 0, 1, \dots, N-1$ are the discrete time-state observation and $N = 30$. Then we simulate the linear dynamical system with the initial state as x_0 and the system matrix A . The relative mean squared error is of $O(1e-3)$. Algorithms 1–4 are performed with the original data from LIP, and we compare the estimated eigenvalues with those of A . The result is presented as below.

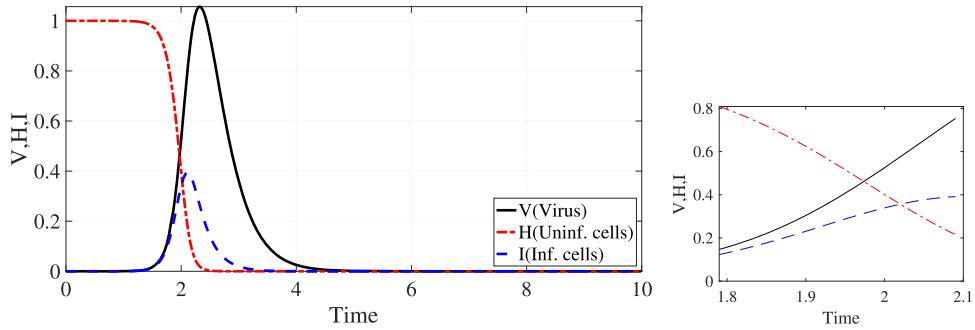


Figure 5. Example 5. Virus(V) infects susceptible cells(H) with rate β . Infected cells are cleared with rate δ . Once cells are productively infected (I), they release virus at rate r and virus are cleared at rate c . The susceptible cells (red line) are rapidly infected while the virus (black line) and infected cells (blue line) peak at $t = 2.2$ approximately. The viral growth is limited by the number of susceptible cells, decreasing the viral load and the number of infected cells to undetectable levels.

Example 6. Human walk motion. In this example, we consider the captured motion of a walking human³. We used the one that is available in Yuying Liu’s github. The data set is of size 54×316 , which is collected by 18 sensors and each sensor provides (x, y, z) information of motion. The total number of time frames is 316 (the gap between each time frame is $1/120$ s). We first normalize the real data set so their mean is 0 and the standard deviation is 1. We then apply the dynamic decomposition method to obtain A . The relative mean square error is $O(10^{-3})$. We use the real motion data in our algorithm and compare the outputs with the eigenvalues of A . We report its result as follows. We see that matrix-pencil-SVD algorithm has the best performance.

Summary. In the case of noise-free data, we have seen that all algorithms provide an accurate reconstruction of the eigenvalues of various affine systems in Λ , matching our theoretical results developed in section 2. They can also provide a faithful approximation of eigenvalues when the time series data is approximately generated by a linear system. For comparison of algorithms, we found that the matrix pencil method has the best performance in most cases when $|\Omega|$ is small and in the real human motion data, demonstrating its robustness. When $|\Omega|$ increases and the data is exact, the Prony method has the best performance in most cases. In particular,

- As $|\Omega|$ increases, more eigenvalues could be recovered and the reconstruction accuracy got improved until $|\Omega|$ is sufficiently large. This is because larger $|\Omega|$ would yield a larger matrix that is sensitive to the perturbation caused by numerical round-off errors.
- For the temporal steps, setting $M = 3d$ is typically enough since the dynamical systems considered in our numerical examples have reached equilibrium (see the state plot in example 1). In these cases, increasing more temporal samples does not bring any new information and did not help in improving the accuracy. An exception case is that the performance of ESPRIT still got improved.

³The data set was obtained from mocap.cs.cmu.edu, which is the first trial of CMU Mocap subject 07 and created with funding from NSF EIA-0196217.

- The difficulty that affects the reconstruction accuracy for large affine systems is caused by the clustering phenomenon of eigenvalues, i.e., the minimal gap between eigenvalues is close to zero. The round-off errors prevented us from recovering eigenvalues that prove in theory, but the numerical example in the sphere graph shows that we can recover significant eigenvalues or representative eigenvalues in a cluster.
- Finally, we have also numerically demonstrated that the algebraic multiplicities of eigenvalues can be recovered when A is similar to a Jordan matrix (see example 1), while we lost this information when A is diagonalizable (see example 3).

5. Conclusion and future work

This paper studied estimating the spectrum of affine systems from partial observations of a single trajectory data. We derived various characterizations on the interplay among the observational locations, the behavior of the dynamical systems, and the spectral properties of the system matrix for the recoverability of eigenvalues. We propose several algorithms, which allow the usage of space–time samples with flexible temporal length and have been applied to a wide variety of examples on both synthetic and real data sets.

Several interesting questions are left for future investigations. First, we would like to address the ‘optimal selection’ of observational locations. Given a fixed number of sensors, some choices of locations perform better than others in terms of numerical stability. We would like to find a characterization. Second, devising denoising algorithm when the observation data is corrupted by noise. One direction we would like to pursue is, when we have multiple observation locations, how to make use of the temporal correlation between them to denoise the Hankel-type matrix. Third, we only consider the recovery of eigenvalues in this paper. It would be interesting to explore when the corresponding eigenspace projections can also be recovered. This problem is related to the completion of the low-rank matrix. We would like to derive conditions under which the recovery is feasible and propose robust algorithms to find a faithful approximation to the original system.

Acknowledgments

The authors would like to thank anonymous reviewers for their insightful comments and constructive feedback that help improve the paper. Sui Tang is supported by Regents Junior Faculty fellowship and Faculty Early Career Acceleration grant sponsored by University of California Santa Barbara and NSF DMS #2111303.

Data availability statement

The data that support the findings of this study will be openly available following an embargo at the following URL/DOI: <https://github.com/jcheng328/DynamicalSampling>. Data will be available from 28 May 2021.

Appendix

Lemma 3. *Let $A = UJU^{-1}$ be its Jordan decomposition as in (10) and $\{V_s\}_{s=1}^n$ be its corresponding invariant subspaces. For any $b \in \mathbb{C}^d$, we have $q_b^A = \prod_{s=1}^n q_{b_s}^A$, where*

$b_s = P(\lambda_s; A)b \in V_s$ with $P(\lambda_s; A)$ the projection onto V_s and the polynomials $\{q_{b_s}^A\}_{s=1}^n$ are coprime with each other.

Proof. One hand, we prove that $\prod_{s=1}^n q_{b_s}^A$ divides q_b^A . Note the annihilating polynomials of A of b_s form an ideal generated by $q_{b_s}^A$:

$$\mathcal{I}_{b_s}^A := \{p(z) \in \mathbb{C}[z] \mid p(A)(b_s) = 0\} = \langle q_{b_s}^A \rangle. \quad (52)$$

Denote by $A|_{V_s}$ the restriction of A on the invariant space V_s and by A_s its matrix form under standard basis. Let $q^{A_s}(t) = (t - \lambda_s)^{r^{A_s}}$ denote the minimal polynomial of A_s and then $q^{A_s}(A_s) \equiv 0$, which implies

$$q^{A_s}(A_s)b_s = 0, \quad (53)$$

so from the ideal (52) property it follows that $q_{b_s}^A$ divides q^{A_s} . So $q_{b_s}^A(z) = (z - \lambda_i)^{r_{b_s}^A}$. Since eigenvalues $\{\lambda_s\}_{s=1}^n$ are distinct, we know that $q_{b_s}^A$ are coprime to each other.

Note that $\sum_{s=1}^n q_b^A(A)b_s = q_b^A(A)b = 0$ and $q_b^A(A)b_s \in V_s$. It follows that $q_b^A(A)b_s = 0$. By the property of ideal (52), we know that $q_{b_s}^A$ divides q_b^A for $s = 1, \dots, n$. Therefore, $\prod_{s=1}^n q_{b_s}^A$ divides q_b^A .

On the other hand, we show that q_b^A divides $\prod_{s=1}^n q_{b_s}^A$. Define

$$\mathcal{I}_b^A = \{p \in \mathbb{C}[z] \mid p(A)b = 0\} = \langle q_b^A \rangle. \quad (54)$$

Note that

$$\prod_{s=1}^n q_{b_s}^A(A)b = \prod_{s=1}^n q_{b_s}^A(A) \sum_{s=1}^n b_s = \sum_{s=1}^n \prod_{s=1}^n q_{b_s}^A(A)b_s = 0, \quad (55)$$

it follows that q_b^A divides $\prod_{s=1}^n q_{b_s}^A$. Hence the conclusion follows. \square

Lemma 4. For $S \in \mathbb{C}^{m \times d}$, we have that $q_{S,b}^A = \prod_{s=1}^n q_{S,b_s}^A$, where q_{S,b_s}^A are coprime with each other. As a result, $r_{S,b}^A = \sum_{s=1}^n r_{S,b_s}^A$.

Proof. First, for each s , we claim that (S, A, b_s) -annihilating polynomial forms an ideal generated by q_{S,b_s}^A , i.e.,

$$\mathcal{I}_{S,b_s}^A = \{p \in \mathbb{C}[z] \mid Sp(A)\mathcal{K}_{r_S^A}(A, b_s) = \{0\}\} = \langle q_{S,b_s}^A \rangle. \quad (56)$$

This is due to the fact (see the proof of lemma 2.3 in [6]) that

$$Sq_{S,b_s}^A(A)\mathcal{K}_{r_S^A}(A, b_s) = \{0\} \iff Sq_{S,b_s}^A(A)h(A)b_s = 0, \quad \text{for any polynomial } h(z).$$

From the definition, it is straightforward to see that q_{S,b_s}^A divides q_b^A . So by lemma 3, we have

$$q_{S,b_s}^A(z) = (z - \lambda_i)^{r_{S,b_s}^A}, \quad (57)$$

and it follows that q_{S,b_s}^A is coprime to q_{S,b_j}^A if $s \neq j$.

Now for each $s = 1, \dots, n$, let $p_s(A) = \prod_{j=1, j \neq s}^n q_{b_j}^A$. Then $p_s(A)b = p_s(A)b_s$. For any polynomial h , we have

$$Sq_{S,b}^A(A)h(A)p_s(A)b_s = Sq_{S,b}^A(A)h(A)p_s(A)b = 0, \quad (58)$$

holds true. So from the property of idea (56), we have q_{S,b_s}^A divides $q_{S,f}^A p_s(A)$. Since q_{S,b_s}^A is coprime to $p_s(A)$, we have q_{S,b_s}^A divides $q_{S,b}^A$. Combining the fact that q_{S,b_s}^A is coprime to q_{S,b_j}^A if $s \neq j$, we have that $\prod_{s=1}^n q_{S,b_s}^A$ divides $q_{S,b}^A$.

On the other hand, for (S, A, b) -annihilating polynomials, we have the ideal

$$\mathcal{I}_{S,b}^A = \{p \in \mathbb{C}[z] \mid Sp(A)\mathcal{K}_{r_S^A}(A, b) = 0\} = \langle q_{S,b}^A \rangle. \quad (59)$$

Note that for any polynomial $h[z]$, we have

$$S \prod_{j=1}^n q_{S,b_j}^A(A)h(A)b = S \prod_{j=1}^n q_{S,b_j}^A(A)h(A) \sum_{s=1}^n b_s = \sum_{s=1}^n Sh(A) \prod_{s=1, s \neq j}^n q_{S,b_s}^A(A)q_{S,b_j}^A(A)b_j = 0, \quad (60)$$

it follows that $q_{S,b}^A$ divides $\prod_{s=1}^n q_{S,b_s}^A$.

Therefore

$$q_{S,b}^A = \prod_{s=1}^n q_{S,b_s}^A. \quad (61)$$

□

Lemma 5. Let $b \in \mathbb{C}^d$, $A \in \mathbb{C}^{d \times d}$ and $S \in \mathbb{C}^{m \times d}$. Given $r_{S,b}^A$, then $q_{S,b}^A$ is the unique monic polynomial q satisfying $\deg(q) \leq r_{S,b}^A$ and the following system of linear equations:

$$Sq(A)A^t b = 0, \quad t = 0, \dots, r_{S,b}^A - 1. \quad (62)$$

Proof. First of all, we claim that the solutions to (62) is the same with the solutions to the system of linear equations:

$$Sq(A)A^t b = 0, \quad t = 0, \dots, r_b^A - 1. \quad (63)$$

Suppose that q is a solution to (62), then for any $j \geq r_{S,b}^A$,

$$A^j = p_j(A)q_{S,b}^A(A) + h_j(A), \quad \deg(h_j(A)) \leq r_{S,b}^A - 1.$$

Therefore,

$$\begin{aligned} Sq(A)A^j b &= Sq(A)p_j(A)q_{S,b}^A(A)b + Sq(A)h_j(A)b \\ &= Sq_{S,b}^A(A)q(A)p_j(A)b + Sq(A)h_j(A)b, \end{aligned} \quad (64)$$

$$= 0 + 0 = 0, \quad (65)$$

where we use the property of $q_{S,b}^A(A)$ in (58) to obtain that $Sq_{S,b}^A(A)q(A)p_j(A)b = 0$. The claim is proved. Therefore, $q = q_{S,b}^A$ by $\deg(q) \leq r_{S,b}^A$ and the definition of $q_{S,b}^A$. \square

Theorem 6 (Theorem 2.6 in [2]). *Let $J \in \mathbb{C}^{d \times d}$ be a matrix in Jordan form as in (10). Let $\{f_i : i = 1, \dots, m\} \subset \mathbb{C}^d$ be a finite subset of vectors, recall that $r_{f_i}^J$ is the degree of the (J, f_i) -annihilator, $l_i = r_i - 1$, and $P_J = \{P_s : s = 1, \dots, n\}$ be the penthouse family for J introduced in definition 1.*

Then the following statements are equivalent:

- (a) *The set of vectors $\{J^j f_i : i = 1, \dots, m, j = 0, \dots, l_i\}$ spans \mathbb{C}^d .*
- (b) *The set of vectors $\{P(\lambda_j; J)N^{l_j} f_i : i = 1, \dots, m, j = 0, \dots, l_i\}$ spans V_j for $j = 1, \dots, n$.*
- (c) *For every $s = 1, \dots, n$, the set $\{P_s f_i, i = 1, \dots, m\}$ spans $E_s = P_s \mathbb{C}^n$.*

ORCID iDs

Sui Tang  <https://orcid.org/0000-0003-3284-5123>

References

- [1] Aldroubi A, Cabrelli C, Çakmak A F, Molter U and Petrosyan A 2017 Iterative actions of normal operators *J. Funct. Anal.* **272** 1121–46
- [2] Aldroubi A, Cabrelli C, Molter U and Tang S 2017 Dynamical sampling *Appl. Comput. Harmon. Anal.* **42** 378–401
- [3] Aldroubi A, Davis J and Krishtal I 2013 Dynamical sampling: time–space trade-off *Appl. Comput. Harmon. Anal.* **34** 95–503
- [4] Aldroubi A, Davis J and Krishtal I 2015 Exact reconstruction of signals in evolutionary systems via spatiotemporal trade-off *J. Fourier Anal. Appl.* **21** 11–31
- [5] Aldroubi A, Huang L X and Petrosyan A 2019 Frames induced by the action of continuous powers of an operator *J. Math. Anal. Appl.* **478** 1059–84
- [6] Aldroubi A and Krishtal I 2014 Krylov subspace methods in dynamical sampling (arXiv:1412.1538)
- [7] Aldroubi A, Krishtal I and Tang S 2020 Phaseless reconstruction from space–time samples *Appl. Comput. Harmon. Anal.* **48** 395–414
- [8] Aldroubi A, Krishtal I and Weber E 2015 Finite dimensional dynamical sampling: an overview *Excursions in Harmonic Analysis* vol 4 (Berlin: Springer) pp 231–44
- [9] Batenkov D and Yomdin Y 2013 On the accuracy of solving confluent Prony systems *SIAM J. Appl. Math.* **73** 134–54
- [10] Beinert R and Hasannasab M 2021 Phase retrieval and system identification in dynamical sampling via Prony’s method (arXiv:2103.10086)
- [11] Beylkin G, Cramer R, Fann G and Harrison R J 2007 Multiresolution separated representations of singular and weakly singular operators *Appl. Comput. Harmon. Anal.* **23** 235–53
- [12] Bonate P L *et al* 2011 *Pharmacokinetic-Pharmacodynamic Modeling and Simulation* vol 20 (Berlin: Springer)
- [13] Bossmann F, Plonka G, Peter T, Nemitz O and Schmitte T 2012 Sparse deconvolution methods for ultrasonic NDT *J. Nondestruct. Eval.* **31** 225–44
- [14] Cabrelli C, Molter U, Paternostro V and Philipp F 2020 Dynamical sampling on finite index sets *J. Anal. Math.* **140** 637–67
- [15] Candès E J, Romberg J and Tao T 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information *IEEE Trans. Inform. Theory* **52** 489–509
- [16] Christensen O and Hasannasab M 2019 Frame properties of systems arising via iterated actions of operators *Appl. Comput. Harmon. Anal.* **46** 664–73
- [17] Chung F R and Graham F C 1997 *Spectral Graph Theory* vol 92 (Providence, RI: American Mathematical Society)

[18] Coutino M, Isufi E, Maehara T and Leus G 2020 State-space network topology identification from partial observations *IEEE Trans. Signal Inf. Process. Netw.* **6** 211–25

[19] Deri J A and Moura J M 2016 New York city taxi analysis with graph signal processing *IEEE Global Conf. Signal and Information Processing (GlobalSIP)* (Piscataway, NJ: IEEE) pp 1275–9

[20] Dong X, Thanou D, Frossard P and Vandergheynst P 2016 Learning Laplacian matrix in smooth graph signal representations *IEEE Trans. Signal Process.* **64** 6160–73

[21] Duan X, Rubin J E and Swigon D 2020 Identification of affine dynamical systems from a single trajectory *Inverse Problems* **36** 085004

[22] Filbir F, Mhaskar H N and Prestin J 2012 On the problem of parameter estimation in exponential sums *Constr. Approx.* **35** 323–43

[23] Godfrey K 1983 *Compartmental Models and Their Application* (London: Academic)

[24] Hanke M 2012 One shot inverse scattering via rational approximation *SIAM J. Imaging Sci.* **5** 465–82

[25] Hespanha J P 2018 *Linear Systems Theory* (Princeton, NJ: Princeton University Press)

[26] Holter N S, Maritan A, Cieplak M, Fedoroff N V and Banavar J R 2001 Dynamic modeling of gene expression data *Proc. Natl Acad. Sci.* **98** 1693–8

[27] Hua Y and Sarkar T K 1990 Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise *IEEE Trans. Acoust. Speech Signal Process.* **38** 814–24

[28] Hua Y and Sarkar T K 1991 On SVD for estimating generalized eigenvalues of singular matrix pencil in noise *IEEE Trans. Signal Process.* **39** 892–900

[29] Ioannidis V N, Romero D and Giannakis G B 2018 Inference of spatio-temporal functions over graphs via multikernel Kriged Kalman filtering *IEEE Trans. Signal Process.* **66** 3228–39

[30] Ioannidis V N, Shen Y and Giannakis G B 2018 Semi-blind inference of topologies and signals over graphs *IEEE Data Science Workshop (DSW)* (IEEE) pp 165–9

[31] Lovász L *et al* 1993 Random walks on graphs: a survey *Combinatorics, Paul Erdos is Eighty* vol 2 pp 1–46

[32] Mateos G, Segarra S, Marques A G and Ribeiro A 2019 Connecting the dots: identifying network structure via graph signal processing *IEEE Signal Process. Mag.* **36** 16–43

[33] Pasdeloup B, Gripon V, Mercier G, Pastor D and Rabbat M G 2016 Characterization and inference of weighted graph topologies from observations of diffused signals (arXiv:1605.02569)

[34] Pasdeloup B, Gripon V, Mercier G, Pastor D and Rabbat M G 2017 Characterization and inference of graph diffusion processes from observations of stationary signals *IEEE Trans. Signal Inf. Process. Netw.* **4** 481–96

[35] Perraudin N, Paratte J, Shuman D, Martin L, Kalofolias V, Vandergheynst P and Hammond D K 2016 Gspbox: a toolbox for signal processing on graphs

[36] Peter T and Plonka G 2013 A generalized Prony method for reconstruction of sparse sums of eigenfunctions of linear operators *Inverse Problems* **29** 025001

[37] Potts D and Tasche M 2010 Parameter estimation for exponential sums by approximate Prony method *Signal Process.* **90** 1631–42

[38] Potts D and Tasche M 2011 Nonlinear approximation by sums of nonincreasing exponentials *Appl. Anal.* **90** 609–26

[39] Rahman M D and Kai-Bor Y 1987 Total least squares approach for frequency estimation using linear prediction *IEEE Trans. Acoust. Speech Signal Process.* **35** 1440–54

[40] Roy R, Paulraj A and Kailath T 1986 ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise *IEEE Trans. Acoust. Speech Signal Process.* **34** 1340–2

[41] Segarra S, Marques A G, Mateos G and Ribeiro A 2017 Network topology inference from spectral templates *IEEE Trans. Signal Inf. Process. Netw.* **3** 467–83

[42] Spielman D A 2007 Spectral graph theory and its applications *48th Annual IEEE Symp. Foundations of Computer Science (FOCS'07)* (Piscataway, NJ: IEEE) pp 29–38

[43] Sporns O 2010 *Networks of the Brain* (Cambridge, MA: MIT Press)

[44] Stanković S, Rubin J E and Swigon D 2014 Identifiability of linear and linear-in-parameters dynamical systems from a single trajectory *SIAM J. Appl. Dyn. Syst.* **13** 1792–815

[45] Steedly W M, Ying C-H J and Moses R L 1994 Statistical analysis of tls-based Prony techniques *Automatica* **30** 115–29

[46] Stevenhagen P and Lenstra H W 1996 Chebotarëv and his density theorem *Math. Intell.* **18** 26–37

[47] Tang S 2017 System identification in dynamical sampling *Adv. Comput. Math.* **43** 555–80

[48] Tang S 2017 Universal spatiotemporal sampling sets for discrete spatially invariant evolution processes *IEEE Trans. Inf. Theory* **63** 5518–28

- [49] Thanou D, Dong X, Kressner D and Frossard P 2017 Learning heat diffusion graphs *IEEE Trans. Signal Inf. Process. Netw.* **3** 484–99
- [50] Viberg M 1995 Subspace-based methods for the identification of linear time-invariant systems *Automatica* **31** 1835–51
- [51] Zhang J and Moura J M F 2014 Diffusion in social networks as sis epidemics: beyond full mixing and complete graphs *IEEE J. Sel. Top. Signal Process.* **8** 537–51