Adv Comput Math (2017) 43:555–580
DOI 10.1007/s10444-016-9497-5

System identification in dynamical sampling

Sui Tang[1]

Received: 30 August 2015 / Accepted: 8 November 2016/
Published online: 22 November 2016
© Springer Science+Business Media New York 2016

Abstract We consider the problem of spatiotemporal sampling in a discrete infinite dimensional spatially invariant evolutionary process $x^{(n)} = A^n x$ to recover an unknown convolution operator A given by a filter $a \in \ell^1(\mathbb{Z})$ and an unknown initial state x modeled as a vector in $\ell^2(\mathbb{Z})$. Traditionally, under appropriate hypotheses, any x can be recovered from its samples on $\mathbb{Z}$ and A can be recovered by the classical techniques of deconvolution. In this paper, we will exploit the spatiotemporal correlation and propose a new sampling scheme to recover A and x that allows us to sample the evolving states $x, Ax, \cdots, A^{N-1}x$ on a sub-lattice of $\mathbb{Z}$, and thus achieve a spatiotemporal trade off. The spatiotemporal trade off is motivated by several industrial applications (Lu and Vetterli, 2249–2252, 2009). Specifically, we show that

$$\{x(m\mathbb{Z}), Ax(m\mathbb{Z}), \cdots, A^{N-1}x(m\mathbb{Z}) : N \geq 2m\}$$

contains enough information to recover a typical "low pass filter" a and x almost surely, thus generalizing the idea of the finite dimensional case in Aldroubi and Krishtal, arXiv:1412.1538 (2014). In particular, we provide an algorithm based on a generalized Prony method for the case when both a and x are of finite impulse response and an upper bound of their support is known. We also perform a perturbation analysis based on the spectral properties of the operator A and initial state x, and verify the results by several numerical experiments. Finally, we provide several other numerical techniques to stabilize the proposed method, with some examples to demonstrate the improvement.

Communicated by: Yang Wang

The research of this work is supported by NSF Grant DMS-1322099

✉ Sui Tang
sui.tang@vanderbilt.edu

[1] Department of Mathematics, Vanderbilt University, Nashville, TN 37240, USA

Keywords Discrete fourier analysis · Distributed sampling · Reconstruction · Channel estimation

Mathematics Subject Classification (2010) Primary 94A20 · 94A12 · 42C15 · 15A29

1 Introduction

1.1 The dynamical sampling problem

In many situations of practical interest, physical systems evolve in time under the action of well-studied operators, an example of which is a diffusion process. Sampling of such an evolving system can be done by sensors or measurement devices that are placed at various locations and can be activated at different times. In practice, increasing the spatial sampling density is usually much more expensive than increasing the temporal sampling rate [34]. Given the different costs associated with spatial and temporal sampling, we aim to reconstruct any states in the evolutionary process using as few sensors as possible, but allow one to take samples at different time levels. In some cases, obtaining samples at a sufficient rate at any single time level may not even be possible; however, spatiotemporal sampling may resolve this issue by oversampling in time. A natural question is whether one can compensate for insufficient spatial sampling densities by oversampling in time. This setting has departed from classical sampling theory, where the samples are taken simultaneously at only one time level, see [3–5, 13, 15–20, 32, 33, 36, 37, 39, 42, 43]. Dynamical sampling is a newly proposed mathematical sampling framework. It involves studying the time-space patterns formed by the locations of the measurement devices and the times of their activation. Mathematically speaking, suppose x is an initial distribution that is evolving in time satisfying the evolution rule:

$$x_t = A_t x$$

where $\{A_t\}_{t\in[0,\infty)}$ is a family of evolution operators satisfying the condition $A_0 = I$. Dynamical sampling asks the question: when do coarse samplings taken at varying times $\{x|_{\Omega_0}, (A_{t_1}x)|_{\Omega_1}, \ldots, (A_{t_N}x)|_{\Omega_N}\}$ contain the same information as a finer sampling taken at the earliest time? One goal of dynamical sampling is to find all spatiotemporal sampling sets $(\chi, \tau) = \{\Omega_t, t \in \tau\}$ such that certain classes of signals x can be recovered from the spatiotemporal samples $x_t(\Omega_t), t \in \tau$. In the above cases, the evolution operators are assumed to be known. It has been well-studied in the context of various evolutionary systems in a very general setting, see [2, 11, 14, 21, 23, 25].

Another important problem arises when the evolution operators are themselves unknown or partially known. In this case, we are interested in finding all spatiotemporal sampling sets and certain classes of evolution operators so that the family $\{A_t\}_{t\in[0,\infty}$ can be identified, or at least its spectrum and unknown states. We call such a problem the unsupervised system identification problem in dynamical sampling.

Dynamical sampling theory finds natural application to the field of Wireless Sensor Networks (WSN). WSN are widely used in many industries, including the health, military, and environmental industries; see [38] for numerious examples. In WSN, a huge amount of sensors are distributed to monitor a physical field such as pollution, temperature or pressure. The goal is to exploit the evolutionary structure and the placement of sensors to reconstruct an unknown field; however, the current approaches and algorithms do not make use of the evolutionary structure of the field, see [26, 30, 31]. Due to physical or economical constraints, it is not always feasible to place sampling devices at the desired locations; however, we may be able to recover the desired information by placing the sensors elsewhere and using evolution process to recover the signals at the relevant locations. In addition, dynamical sampling will make the reconstruction cheaper since we use a reduced numer of sensors.

1.2 Problem statement

In this subsection, we state an instance of the unsupervised system identification problem of dynamical sampling in an infinite dimensional setting. Let $x \in \ell^2(\mathbb{Z})$ be an unknown initial spatial signal and the evolution operator A be given by an unknown convolution filter $a \in \ell^1(\mathbb{Z})$ such that $Ax = a * x$. At time $t = n \in \mathbb{N}$, the signal x evolves to be $x_n = A^n x = a^n * x$, where $a^n = a * a \cdots * a$. We call this evolutionary system *spatially invariant*. Given some set of spatiotemporal samples with both x and A unknown, we would like to recover as much information as possible given the prior assumptions. Here, we first study the case of uniform subsampling. Without loss of generality, we assume a positive odd integer m ($m > 1$) to be the uniform subsampling factor. At time level $t = l$, we uniformly undersample the evolving state $A^l x$ and get the spatiotemporal data

$$y_l = (a^l * x)(m\mathbb{Z}), \tag{1.1}$$

which is a sequence in $\ell^2(\mathbb{Z})$. It is obvious that at any single time level $t = l$, we cannot determine the state $A^l x$ from the measurement y_l. The problem we are going to consider can be summarized as follows:

Under what conditions on a, m, N and x, can a and x be recovered from the spatiotemporal samples $\{y_l : l = 0, \cdots, N-1\}$, or equivalently, from the set of measurement sequences $\{x(m\mathbb{Z}), (a * x)(m\mathbb{Z}), \cdots, (a^{N-1} * x)(m\mathbb{Z})\}$?

In [1], Aldroubi and Krishtal consider the recovery of an unknown $d \times d$ matrix B and an unknown initial state $x \in \ell^2(\mathbb{Z}_d)$ from coarse spatial samples of its successive states $\{B^k x, k = 0, 1, \cdots\}$. Given an initial sampling set $\Omega \subset \mathbb{Z}_d = \{1, 2, \cdots, d\}$, they employ techniques related to Krylov subspace methods to show how large l_i should be to recover all the eigenvalues of B that can possibly be recovered from spatiotemporal samples $\{B^k x(i) : i \in \Omega, k = 0, 1, \cdots, l_i - 1\}$. Our setup is very similar to the special case of the regular invariant dynamical sampling problem in [1]. In this special case, they employ a generalization of the well known Prony method that uses these regular undersampled spatiotemporal data first for the recovery of the filter a. Then by using techniques developed in [21], they show how to recover the initial state from these spatiotemporal samples. In this paper, we will address

the infinite dimensional analog of this special case and provide more algorithms. In [12], Peter and Plonka use a generalized Prony method to reconstruct the sparse sums of the eigenfunctions of some known linear operators. Our generalization of Prony method shares some similar spirits with that of [12] , but deals with a fundamentally different problem. In Sparse Fourier Transform, see [7–10], the idea is to uniformly undersample a fixed signal with different factors so that one can group subsets of Fourier space together into a small number of bins to isolate frequencies, then take an Aliasing-Based Search by Chinese Remainder Theorem so that one can recover the coefficients and the frequencies. In our case, intuitively, one can think of recovering the shape of an evolving wave by observing the amplitude of its aliased version at fixed coarse locations over a long period of time as opposed to acquiring all of the amplitudes at once. Then by the given priors, one can achieve the perfect reconstructions. Other similar work includes the the Slepian-Wolf distributed source coding problem [6] and the distributed sampling problem in [35]. Our problem, however, is very different from the above in the nature of the processes we study. The distributed sampling problem typically deals with two signals correlated by a transmission channel. We, on the other hand, can observe an evolution process at several instances and over longer periods of time.

1.3 Notation

In the following, we use standard notations. By $\mathbb{N}$, we denote the set of all positive integers. 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 $\mathbf{A}_{M,N} = (a_{ij}) \in \mathbb{C}^{M\times N}$, its transpose is denoted by $\mathbf{A}^T_{M,N}$, its conjugate-transpose by $\mathbf{A}^*_{M,N}$,and its Moore-Penrose pseudoinverse by $\mathbf{A}^+_{M,N}$. A square matrix $\mathbf{A}_{M,M}$ is abbreviated to $\mathbf{A}_M$. Its infinity norm is defined by

$$||\mathbf{A}_M||_\infty = \max_{1\le i\le M}(\sum_{j=1}^{M}|a_{ij}|).$$

For a vector $\mathbf{z} = (z_i) \in \mathbb{C}^M$, the $M \times M$ diagonal matrix built from $\mathbf{z}$ is denoted by $diag(\mathbf{z})$. We define the infinity norm $||\mathbf{z}||_\infty = \max_{i=1,\cdots,M}|z_i|$. It is easy to see that

$$||\mathbf{A}_M||_\infty = \max_{\mathbf{z}\in\mathbb{C}^M,||\mathbf{z}||_\infty=1}||\mathbf{A}_M\mathbf{z}||_\infty.$$

Further, we use submatrix notation similar to that of MATLAB. For example, $\mathbf{A}_{M,M+1}(1 : M, 2 : M + 1)$ is the submatrix of $\mathbf{A}_{M,M+1}$ obtained by extracting rows 1 through M and columns 2 through $M + 1$, and $\mathbf{A}_{M,M+1}(1 : M, M + 1)$ means the last column vector of $\mathbf{A}_{M,M+1}$.

Definition 1.1 The minimal annihilating polynomial of a square matrix $\mathbf{A}_M$ is $p^{\mathbf{A}_M}[z]$, and is the monic polynomial of smallest degree among all the monic polynomials p such that $p(\mathbf{A}_M) = 0$. We will denote the degree of $p^{\mathbf{A}_M}[z]$ by $deg(p^{\mathbf{A}_M})$.

Given a monic polynomial $p[z] = \sum_{k=0}^{M-1} p_k z^k + z^M$, the companion matrix of $p[z]$ is defined by

$$\mathbf{C}^{p[z]} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -p_0 \\ 1 & 0 & \cdots & 0 & -p_1 \\ 0 & 1 & \cdots & 0 & -p_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -p_{M-1} \end{pmatrix}.$$

Definition 1.2 Let $w_1, w_2, \cdots, w_n$ be n distinct complex numbers, and denote $\mathbf{w} = [w_1, \cdots, w_n]^T$. Then the $n \times N$ Vandermonde matrix generated by $\mathbf{w}$ is defined by

$$\mathbf{V}_{n,N}(\mathbf{w}) = \begin{pmatrix} 1 & w_1 & \cdots & w_1^{N-1} \\ 1 & w_2 & \cdots & w_2^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w_n & \cdots & w_n^{N-1} \end{pmatrix}. \tag{1.2}$$

Definition 1.3 For a sequence $c = (c_n)_{n\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$ or $\ell^2(\mathbb{Z})$, we define its Fourier transform to be the function on the Torus $\mathbb{T} = [0, 1)$

$$\hat{c}(\xi) = \sum_{n\in\mathbb{Z}} c_n e^{-2\pi i n\xi}, \xi \in \mathbb{T}.$$

The remainder of the paper is organized as follows: In Section 2, we discuss the noise-free case and show that we can reconstruct a typical low pass filter a and the initial signal x from the dynamical spatiotemporal samples $\{y_l\}_{l=0}^{N-1}$ almost surely, provided $N \geq 2m$. For the case when both a and x are of finite impulse response and an upper bound of their support is known, we propose a Generalized Prony Method algorithm to recover the Fourier spectrum of a. In Section 3, we provide a perturbation analysis of this algorithm. The estimation results are formulated in the rigid ℓ^∞ norm and give us an idea of how the performance depends on the system parameters a, x and m. In Section 4, we do several numerical experiments to verify the results from Section 3. In Section 5, we propose several other algorithms such as Generalized Matrix Pencil , Generalized ESPRIT and Cadzow Denoising methods to improve the effectiveness and robustness of recovery. The comparison between algorithms is illustrated by several numerical examples in Section 6. Finally, we summarize the work in section 7.

2 Noise-free recovery

We consider the recovery of a frequently encountered case in applications when the filter $a \in \ell^1(\mathbb{Z})$ is a typical low pass filter, i.e. $\hat{a}(\xi)$ is real, symmetric and strictly decreasing on $[0, \frac{1}{2}]$. An example of such a typical low pass filter is shown in Fig. 1.

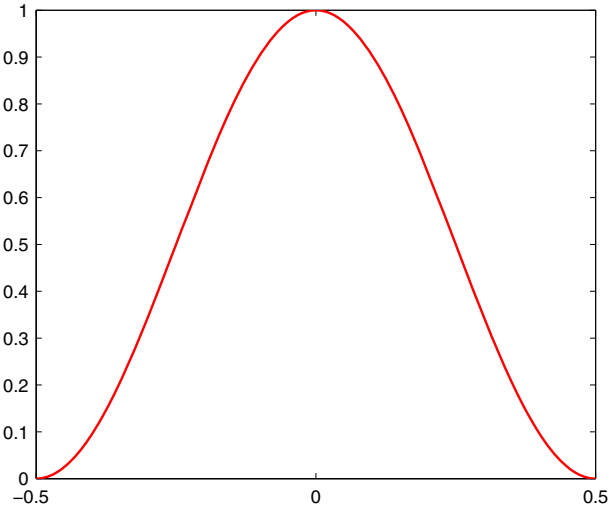

Fig. 1 A Typical Low Pass Filter

The symmetry reflects the fact that there is often no preferential direction for physical kernels and monotonicity is a reflection of energy dissipation.

Without loss of generality, we also assume a is a normalized filter, i.e. $|\hat{a}(\xi)| \leq 1, \hat{a}(0) = 1$. In this section, we assume the spatiotemporal data $y_l = (a^l * x)(m\mathbb{Z})$ is measured exactly. Define the downsampling operator $S_m : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ by

$$(S_m x)(k) = x(mk), k \in \mathbb{Z},$$

then $y_l = S_m(a^l * x)$. Due to the Poisson Summation formula and the convolution theorem, we have the lemma below for the downsampling operator.

Lemma 2.1 *The Fourier transform of each measurement sequence* $y_l = S_m(a^l * x)$ *at* $\xi \in \mathbb{T}$ *is*

$$\hat{y}_l(\xi) = \frac{1}{m} \sum_{i=0}^{m-1} \hat{a}^l(\frac{\xi + i}{m}) \hat{x}(\frac{\xi + i}{m}). \tag{2.1}$$

Let N be an integer satisfying $N \geq 2m$; we define the $(N - m) \times 1$ column vector

$$\mathbf{h}_t(\xi) = [\hat{y}_t(\xi), \hat{y}_{t+1}(\xi), \cdots, \hat{y}_{N-m+t-1}(\xi)]^T, \tag{2.2}$$

and build the Hankel matrices

$$\mathbf{H}^{\xi}_{N-m,m}(0) = \big[\mathbf{h}_0(\xi), \mathbf{h}_1(\xi), \cdots, \mathbf{h}_{m-1}(\xi)\big], \tag{2.3}$$
$$\mathbf{H}^{\xi}_{N-m,m}(1) = \big[\mathbf{h}_1(\xi), \mathbf{h}_2(\xi), \cdots, \mathbf{h}_m(\xi)\big].$$

For $\xi \in \mathbb{T}$, we introduce the notations $\mathbf{x}(\xi) = [\hat{x}(\frac{\xi}{m}), \cdots, \hat{x}(\frac{\xi+m-1}{m})]^T$ and $\mathbf{w}(\xi) = [\hat{a}(\frac{\xi}{m}), \cdots, \hat{a}(\frac{\xi+m-1}{m})]^T$.

Proposition 2.1 *Let N be an integer satisfying $N \geq 2m$.*

(1) *Then the rectangular Hankel matrices can be factorized in the following form:*

$$m\mathbf{H}^{\xi}_{N-m,m}(s) = \mathbf{V}^{T}_{m,N-m}(\mathbf{w}(\xi))diag(\mathbf{x}(\xi))diag(\mathbf{w}(\xi))^{s}\mathbf{V}_{m}(\mathbf{w}(\xi)), \quad (2.4)$$

where $s = 0, 1$. The Vandermonde matrix $\mathbf{V}_{m,N-m}(\mathbf{w}(\xi))$ and $\mathbf{V}_{m}(\mathbf{w}(\xi))$ are given as in Definition 1.2.

(2) *Assume the entries of $\mathbf{x}(\xi)$ are all nonzero. The rank of the Hankel matrix $\mathbf{H}^{\xi}_{N-m,m}(0)$ can be summarized as follows:*

$$Rank\ \mathbf{H}^{\xi}_{N-m,m}(0) = \begin{cases} m & if\ \xi \neq 0\ or \frac{1}{2}, \\ \frac{m+1}{2} & otherwise. \end{cases}$$

(3) *Assume the entries of $\mathbf{x}(\xi)$ are all nonzero. For $\xi \neq 0, \frac{1}{2}$, the vector $\mathbf{q}(\xi) = [q_0(\xi), \cdots, q_{m-1}(\xi)]^T$ is the unique solution of the linear system*

$$\mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{q}(\xi) = -\mathbf{h}_{m}(\xi) \quad (2.5)$$

if and only if the polynomial

$$q^{\xi}[z] = \sum_{k=0}^{m-1} q_k(\xi)z^k + z^m \quad (2.6)$$

with coefficients given by $\mathbf{q}(\xi)$ is the minimal annihilating polynomial of the diagonal matrix $diag(\mathbf{w}(\xi))$. In other words, the polynomial $q^{\xi}[z]$ has all $\hat{a}(\frac{\xi+i}{m}) \in \mathbb{R} (i = 0, \cdots, m-1)$ as roots. Moreover, if $p[z]$ is a monic polynomial of degree m, then

$$\mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{C}^{p[z]} = \mathbf{H}^{\xi}_{N-m,m}(1) \quad (2.7)$$

if and only if $p[z]$ is the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$.

Proof (1) By Lemma 2.1, for $t = 0, \cdots, m-1$, we have the identity:

$$m\mathbf{h}_{t}(\xi) = \mathbf{V}^{T}_{m,N-m}(\mathbf{w}(\xi))diag(\mathbf{x}(\xi))\mathbf{V}_{m}(\mathbf{w}(\xi))(:, t+1).$$

Hence the first identity follows by the definition of $\mathbf{H}^{\xi}_{N-m,m}(0)$. Noticing that for $t \geq 1$,

$$m\mathbf{h}_{t}(\xi) = \mathbf{V}^{T}_{m,N-m}(\mathbf{w}(\xi))diag(\mathbf{x}(\xi))diag(\mathbf{w}(\xi))\mathbf{V}_{m}(\mathbf{w}(\xi))(:, t),$$

the second identity follows similarly.

(2) By the symmetric and monotonicity condition of $\hat{a}$ on $\mathbb{T}$, we have

$$\text{Rank}V_{m}(\mathbf{w}(\xi)) = \begin{cases} m & \text{if } \xi \neq 0 \text{ or } \frac{1}{2}, \\ \frac{m+1}{2} & \text{otherwise.} \end{cases} \quad (2.8)$$

Since $N \geq 2m$, $\text{Rank}\mathbf{V}_{m}(\mathbf{w}(\xi)) = \text{Rank } \mathbf{V}^{T}_{m,N-m}(\mathbf{w}(\xi))$. By our assumptions, $diag(\mathbf{x}(\xi))$ is invertible. The rank of Hankel matrix $\mathbf{H}^{\xi}_{N-m,m}(0)$ can be computed by its factorization in (1).

(3) If $\xi \neq 0$ or $\frac{1}{2}$, then the diagonal matrix $diag(\mathbf{w}(\xi))$ has m distinct eigenvalues consisted of $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$, and therefore the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$ is of degree m. Supposing $q^{\xi}[z] = \sum_{k=0}^{m-1} q_k(\xi) z^k + z^m$ is the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$, $q^{\xi}[diag(\mathbf{w}(\xi))] = 0$. In other words,

$$\sum_{k=0}^{m-1} q_k(\xi) diag(\mathbf{w}(\xi))^k = -diag(\mathbf{w}(\xi))^m.$$

Then

$$\begin{aligned} \mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{q}(\xi) &= \sum_{k=0}^{m-1} q_k(\xi)\mathbf{h}_k(\xi) \\ &= \mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi))(\sum_{k=0}^{m-1} q_k(\xi) diag(\mathbf{w}(\xi))^k)\mathbf{x}(\xi) \\ &= -\mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi)) diag(\mathbf{w}(\xi))^m \mathbf{x}(\xi) \\ &= -\mathbf{h}_m(\xi). \end{aligned} \tag{2.9}$$

Conversely, if $\mathbf{q}(\xi)$ is the solution of linear system (2.5), let the monic polynomial given by $\mathbf{q}(\xi)$ be $q^{\xi}[z]$; then by the computation process of Eq. 2.9, we have

$$\mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi)) q^{\xi}[diag(\mathbf{w}(\xi))]\mathbf{x}(\xi) = 0.$$

Since $\mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi))$ is full column rank, $q^{\xi}[diag(\mathbf{w}(\xi))]\mathbf{x}(\xi) = 0$. By the fact that $q^{\xi}[diag(\mathbf{w}(\xi))]$ is diagonal and $\mathbf{x}(\xi)$ has no zero entries, we know $q^{\xi}[z]$ is a monic annihilating polynomial of $diag(\mathbf{w}(\xi))$. Minimality follows by counting its degree. If $p[z]$ is a monic annihilating polynomial of $diag(\mathbf{w}(\xi))$, it is easy to show (2.7) is an equivalent formulation to the identity (2.5).

□

Corollary 2.1 *In the case $\xi = 0$ or $\frac{1}{2}$, if $diag(\mathbf{x}(\xi))$ is invertible, then the coefficient vector of the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$ $\mathbf{c}(\xi) \in \mathbb{R}^{\frac{m+1}{2}}$ is the unique solution of the following linear system:*

$$\mathbf{H}^{\xi}_{N-m,\frac{m+1}{2}}(0)\mathbf{c}(\xi) = -\mathbf{h}_{\frac{m+1}{2}}(\xi), \tag{2.10}$$

where $\mathbf{H}^{\xi}_{N-m,\frac{m+1}{2}}(0) = \left[\mathbf{h}_0(\xi), \cdots, \mathbf{h}_{\frac{m-1}{2}}(\xi)\right]$.

Let μ denote the Lebesgue measure on $\mathbb{T}$, and X be a subclass of $\ell^2(\mathbb{Z})$ defined by

$$X = \{x \in \ell^2(\mathbb{Z}) : \mu(\{\xi \in \mathbb{T} : \hat{x}(\xi) = 0\}) = 0\}.$$

Clearly, X is a dense class of $\ell^2(\mathbb{Z})$ under the norm topology. In noise-free scenario, we show that we can recover a and x provided that our initial state $x \in X$.

Theorem 2.2 *Let $x \in X$ be the initial state and the evolution operator A be a convolution operator given by $a \in \ell^1(\mathbb{Z})$ so that $\hat{a}(\xi)$ is real, symmetric, and strictly decreasing on $[0, \frac{1}{2}]$. Then a and x can be recovered from the set of measurement sequences $\{y_l = (a^l * x)(m\mathbb{Z}) : l = 0, \cdots, N-1\}$ defined in (1.1) when $N \geq 2m$.*

Proof Since Fourier transform is an isometric isomorphism from $\ell^2(\mathbb{Z})$ to $L^2(\mathbb{T})$, we can look at this recovery problem on the Fourier domain equivalently. We are going to show that the regular subsampled data $\{y_l\}_{l=0}^{N-1}$ contains enough information to recover the Fourier spectrum of a on $\mathbb{T}$ up to a measure zero set. By our assumptions on x, there exists a measurable subset E_0 of $\mathbb{T}$ with $\mu(E_0) = 1$, so that $diag(\mathbf{x}(\xi))$ is an invertible matrix for $\xi \in E_0$. Let $E = E_0 - \{0, \frac{1}{2}\}$. Then if $\xi \in E$, we can recover the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$ on account of Proposition 2.1. Now to recover the diagonal entries of $diag(\mathbf{w}(\xi))$, it amounts to finding the roots of this minimal annihilating polynomial and ordering them according to the monotonicity and symmetry condition on $\hat{a}$. In summary, for each $\xi \in E$, we can uniquely determine $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$. Note $\mu(E) = 1$, and hence we can recover the Fourier spectrum of a up to a measure zero set. The conclusion is followed by applying the inverse Fourier transform on $\hat{a}(\xi)$. Once a is recovered, we can recover x from the spatiotemporal samples $\{y_l\}_{l=0}^{m-1}$ using techniques developed in [11]. □

Theorem 2.2 addresses the infinite dimensional analog of Theorem 4.1 in [1]. If we don't know anything about a in advance, with minor modifications of the above proof, one can show the recovery of the range of $\hat{a}$ on a measurable subset of $\mathbb{T}$, whose measure is 1.

Definition 2.3 Let $a = (a(n))_{n\in\mathbb{Z}}$, the support set of a is defined by $Supp(a) = \{k \in \mathbb{Z} : a(k) \neq 0\}$. If $Supp(a)$ is a finite set, a is said to be of finite impulse response.

In particular, if x is of finite impulse response, then $x \in X$. Now if both x and a are of finite impulse response, and we know an upper bound $r \in \mathbb{N}$ such that $Supp(a)$ and $Supp(x)$ are contained in $\{-r, -r+1, \cdots, r\}$, then we can compute the value of the Fourier transformation of $\{y_l\}_{l=0}^{N-1}$ at any $\xi \in \mathbb{T}$. From the proof of Theorem 2.2, we can give an algorithm similar to the classical Prony method to recover $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ almost surely, given ξ chosen uniformly from $\mathbb{T}$. This process is summarized in Algorithm 1.

Algorithm 1 Generalized prony method

Input: $N \geq \quad m \;\; r \in \mathbb{N}, \{y_l\}_{l=0}^{N-1} \;\; \xi(\neq 0, \frac{1}{2}) \in \mathbb{T}$.

1: Compute the Fourier transform of the measurement sequences $\{y_l\}_{l=0}^{N-1}$ and build the Hankel matrix $\mathbf{H}^{\xi}_{N-m,m}(0)$ and the vector $\mathbf{h}_m(\xi)$.

2: Compute the solution of the overdetermined linear system (2.5):

$$\mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{q}(\xi) = -\mathbf{h}_m(\xi).$$

Form the polynomial $q^{\xi}[z] = \sum_{k=0}^{m-1} q_k(\xi) z^k + z^m$ and find its roots; this can be done by solving the standard eigenvalue problem of its companion matrix.

3: Order the roots by the monotonicity and symmetry condition of $\hat{a}$ to get $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

Output: $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

Corollary 2.2 *In addition to the assumptions of Theorem 2.2, if both a and x are of finite impulse response with support contained in* $\{-r, -r+1, \cdots, r\}$ *for some* $r \in \mathbb{N}$*, then it is enough to determine a and x after we recover* $\{\hat{a}(\eta_i) : i = 1, \cdots, r\}$ *at r distinct locations by Algorithm 1.*

Proof Under these assumptions, we know

$$\hat{a}(\xi) = a(0) + \sum_{k=1}^{r} a(k)\cos(2\pi k\xi). \tag{2.11}$$

Supposing $\{\hat{a}(\eta_i) : i = 1, \cdots, r, \eta_i \neq \eta_j \text{if } i \neq j\}$ are recovered, we set up the following linear equation

$$\begin{pmatrix} 1 & cos(2\pi\eta_1) & \cdots & cos(2r\pi\eta_1) \\ 1 & cos(2\pi\eta_2) & \cdots & cos(2r\pi\eta_2) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & cos(2\pi\eta_r) & \cdots & cos(2r\pi\eta_r) \end{pmatrix} \begin{pmatrix} a(0) \\ a(1) \\ \vdots \\ a(r) \end{pmatrix} = \begin{pmatrix} \hat{a}(\eta_1) \\ \hat{a}(\eta_2) \\ \vdots \\ \hat{a}(\eta_r) \end{pmatrix}. \tag{2.12}$$

Note that $\{1, \cos(2\pi\eta), \cdots, \cos(2r\pi\eta)\}$ is a Chebyshev system on $[0, 1]$ (see [45]), and hence (4) has a unique solution. Then we can recover x by solving the linear system

$$\mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi))\mathbf{x}(\xi) = \mathbf{h}_0(\xi)$$

for finitely many ξ, which finishes the proof. □

3 Perturbation analysis

In previous sections, we have shown that if we are able to compute the spectral data $\{\hat{y}_l(\xi)\}_{l=0}^{N-1}$ at ξ, then we can recover the Fourier spectrum $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ by Algorithm 1. However, we assume the spectral data are noise-free.

A critical issue still remains: we need to analyze the accuracy of the solution achieved by Algorithm 1 in the presence of noise. Mathematically speaking, assume the measurements are given by $\{\tilde{y}_l\}_{l=0}^{N-1}$ so that $||\hat{y}_l(\xi) - \widehat{\tilde{y}_l}(\xi)||_\infty \leq \epsilon_l$ for all $\xi \in \mathbb{T}$. Given $\epsilon = max_l|\epsilon_l|$, how large can the error be in the worst case for the output parameters of Algorithm 1 in terms of ϵ and the system parameters a, x and m. Most importantly, we need to understand analytically what kind of effects that the subsampling factor m will impose on the performance of Algorithm 1.

In this section, for simplicity, we choose $N = 2m$ to meet the minimal requirement. In this case, the Hankel matrix $\mathbf{H}^{\xi}_{N-m,m}(0)$ is a square matrix and the vectors $\mathbf{h}_t(\xi)$ are of length m. We denote them by two new notations: $\mathbf{H}_m(\xi)$ and $\mathbf{b}_t(\xi)$. Our perturbation analysis will consist of two steps. Suppose our measurements are perturbed from $\{y_l\}_{l=0}^{2m-1}$ to $\{\tilde{y}_l\}_{l=0}^{2m-1}$. For any ξ, we first measure the perturbation of $\mathbf{q}(\xi)$ in terms of the ℓ^∞ norm, then we measure the perturbation of the roots. It is well known that the roots of a polynomial are dependent on the small change in the coefficients (see Proposition 3.1). Hence, for a small perturbation, although the roots of the perturbed polynomial $\tilde{q}^{\xi}[z]$ may not be real, we can order them according to their modulus and have a one to one correspondence with the roots of $q^{\xi}[z]$. Before presenting our main results in this section, let us introduce some useful notations and terminologies.

Definition 3.1 Let $\xi \in \mathbb{T} - \{0, \frac{1}{2}\}$, and consider the set $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ consisting of m distinct nodes.

(1) For $0 \leq k \leq m-1$, the separation between $\hat{a}(\frac{\xi+k}{m})$ with the other $m-1$ nodes is measured by

$$\delta_k(\xi) = \frac{1}{\prod\limits_{\substack{j\neq k \\ 0\leq j\leq m-1}} |\hat{a}(\frac{\xi+j}{m}) - \hat{a}(\frac{\xi+k}{m})|}.$$

(2) For $0 \leq k \leq m$, the k-th elementary symmetric function generated by the m nodes is denoted by

$$\sigma_k(\xi) = \begin{cases} 1 & \text{if } k\text{=0,} \\ \sum\limits_{0\leq j_1<\cdots<j_k\leq m-1} \hat{a}(\frac{\xi+j_1}{m})\hat{a}(\frac{\xi+j_2}{m})\cdots\hat{a}(\frac{\xi+j_k}{m}) & \text{otherwise.} \end{cases} \tag{3.1}$$

For $0 \leq k, i \leq m-1$, the k-th elementary symmetric function generated by $m-1$ nodes with $\hat{a}(\frac{\xi+i}{m})$ missing is denoted by $\sigma_k^{(i)}(\xi)$.

The following proposition measures the perturbation of the polynomial roots in terms of the perturbation of its coefficients and is the key to our perturbation analysis.

Proposition 3.1 (see Proposition **V**.1 in [40]) *Let z_k be a root of multiplicity $M_k \in \mathbb{N}^+$ of the r-th order polynomial $p[z]$. For all $\epsilon > 0$, let $p_\epsilon[z] = p[z]+\epsilon\Delta p[z]$, where $\Delta p[z]$ is a polynomial of order lower than r. Suppose that $\Delta p[z_k] \neq 0$. Then there*

exists a positive ϵ_0 such that for all $\epsilon < \epsilon_0$ there are exactly M_k roots of $p_\epsilon[z]$, denoted $\{z_{k,m}(\epsilon)\}_{m\in\{0,\cdots,M_k-1\}}$, which admit the first-order fractional expansion

$$z_{k,m}(\epsilon) = z_k + \epsilon^{\frac{1}{M_k}} \Delta z_k e^{2\pi i \frac{m}{M_k}} + O(\epsilon^{\frac{2}{M_k}}), \tag{3.2}$$

where Δz_k is an arbitrary M_k-th root of the complex number

$$(\Delta z_k)^{M_k} = -\frac{\Delta p[z_k]}{\frac{1}{M_k!} p^{(M_k)}[z_k]}. \tag{3.3}$$

Proposition 3.2 *Let the perturbed measurements $\{\tilde{y}_l\}_{l=0}^{2m-1}$ be given with an error satisfying $||\widehat{\tilde{y}}_l(\xi) - \hat{y}_l(\xi)||_\infty \le \epsilon, \forall l$. Let $\widetilde{\mathbf{H}}_m(\xi)$ and $\tilde{\mathbf{b}}_m(\xi)$ be given by $\{\widehat{\tilde{y}}_l(\xi)\}_{l=0}^{2m-1}$ in the same way as in (2.3) and (2.1). Assume $\mathbf{H}_m(\xi)$ is invertible and ϵ is sufficient small so that $\widetilde{\mathbf{H}}_m(\xi)$ is also invertible. Denote by $\tilde{\mathbf{q}}(\xi)$ the solution of the linear system $\widetilde{\mathbf{H}}_m(\xi)\tilde{\mathbf{q}}(\xi) = -\tilde{\mathbf{b}}_m(\xi)$. Let $\tilde{q}^\xi[z]$ be the Prony polynomial formed by $\tilde{\mathbf{q}}(\xi)$ and $\{\tilde{\hat{a}}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ be its roots, then we have the following estimates as $\epsilon \to 0$:*

$$||\mathbf{q}(\xi) - \tilde{\mathbf{q}}(\xi)||_\infty \le ||\mathbf{H}_m^{-1}(\xi)||_\infty (1 + m\beta_1(\xi))\epsilon + O(\epsilon^2), \tag{3.4}$$

where $\beta_1(\xi) = \max\limits_{k=1,\cdots,m} |\sigma_k(\xi)|$. As a result, we achieve the following first order estimation

$$|\tilde{\hat{a}}(\frac{\xi+i}{m}) - \hat{a}(\frac{\xi+i}{m})| \le C_i(\xi)(1 + m\beta_1(\xi))||\mathbf{H}_m^{-1}(\xi)||_\infty \epsilon + O(\epsilon^2), \tag{3.5}$$

where $C_i(\xi) = \delta_i(\xi) \cdot (\sum\limits_{k=0}^{m-1} |\hat{a}^k(\frac{\xi+i}{m})|)$.

Proof Note that the linear system (2.5) is perturbed to be

$$\widetilde{\mathbf{H}}_m(\xi)\tilde{\boldsymbol{q}}(\xi) = -\tilde{\mathbf{b}}_m(\xi). \tag{3.6}$$

By our assumptions, we have

$$||\Delta\mathbf{H}_m(\xi)||_\infty = ||\widetilde{\mathbf{H}}_m(\xi) - \mathbf{H}_m(\xi)||_\infty \le m\epsilon, \tag{3.7}$$

$$||\Delta\mathbf{b}_m(\xi)||_\infty = ||\tilde{\mathbf{b}}_m(\xi) - \mathbf{b}_m(\xi)||_\infty \le \epsilon. \tag{3.8}$$

Define $\Delta\mathbf{q}(\xi) = \tilde{\mathbf{q}}(\xi) - \mathbf{q}(\xi)$. Then by simple computation,

$$\Delta\mathbf{q}(\xi) = \mathbf{H}_m^{-1}(\xi)(I + \mathbf{H}_m^{-1}(\xi)\Delta\mathbf{H}_m(\xi))^{-1}(-\Delta\mathbf{b}_m(\xi) - \Delta\mathbf{H}_m(\xi)\mathbf{q}(\xi)). \tag{3.9}$$

Hence if $\epsilon \to 0$, we obtain

$$\Delta\mathbf{q}(\xi) = \mathbf{H}_m^{-1}(\xi)(-\Delta\mathbf{b}_m(\xi) - \Delta\mathbf{H}_m(\xi)\mathbf{q}(\xi)) + O(\epsilon^2). \tag{3.10}$$

Now we can easily get an estimation of ℓ^∞ norm of $\Delta\mathbf{q}(\xi)$

$$||\Delta\mathbf{q}(\xi)||_\infty \le ||\mathbf{H}_m^{-1}(\xi)||_\infty (1 + m||\mathbf{q}(\xi)||_\infty)\epsilon + O(\epsilon^2). \tag{3.11}$$

Since $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ are the roots of $q^{\xi}[z]$, using Vieta's Formulas (see [41]), we know

$$||\mathbf{q}(\xi)||_{\infty} = \max_{1\leq k\leq m} |\sigma_k(\xi)|.$$

Let $(\Delta q(\xi))[z]$ be the polynomial of degree less than or equal to $m-1$ defined by the vector $\Delta\mathbf{q}(\xi)$. Using Proposition 3.1, and denoting by $(q^{\xi})'[z]$ the derivative function of $q^{\xi}[z]$, for $0 \leq i \leq m-1$, we conclude

$$\begin{aligned}
|\tilde{\hat{a}}(\tfrac{\xi+i}{m}) - \hat{a}(\tfrac{\xi+i}{m})| &= |\tfrac{\Delta\mathbf{q}(\xi)[\hat{a}(\frac{\xi+i}{m})]}{(q^{\xi})'[\hat{a}(\frac{\xi+i}{m})]} + O(\epsilon^2)| \\
&\leq \frac{||\Delta\mathbf{q}(\xi)||_{\infty}(\sum_{k=0}^{m-1}|\hat{a}^k(\frac{\xi+i}{m})|)}{\prod_{\substack{j\neq i \\ 0\leq j\leq m-1}}|\hat{a}(\frac{\xi+j}{m})-\hat{a}(\frac{\xi+i}{m})|} + O(\epsilon^2) \\
&\leq C_i(\xi)||\mathbf{H}_m^{-1}(\xi)||_{\infty}(1 + m\, max_{1\leq k\leq m}|\sigma_k(\xi)|)\epsilon + O(\epsilon^2),
\end{aligned} \tag{3.12}$$

where $C_i(\xi) = \delta_i(\xi)(\sum_{k=0}^{m-1}|\hat{a}^k(\frac{\xi+i}{m})|)$. □

Therefore it is important to understand the relation between the behavior of $||\mathbf{H}_m^{-1}(\xi)||_{\infty}$ and our system parameters a, m and x. Consequently, we estimate $||\mathbf{H}_m^{-1}(\xi)||_{\infty}$ and reveal their connection with the spectral properties of a, x and the subsampling factor m.

Proposition 3.3 *Assume $\mathbf{H}_m(\xi)$ is invertible, then we have the lower bound:*

$$||\mathbf{H}_m^{-1}(\xi)||_{\infty} \geq m \cdot \max_{i=0,\cdots,m-1} \frac{\beta_2(i,\xi)\delta_i(\xi)}{|\hat{x}(\frac{\xi+i}{m})|}, \tag{3.13}$$

where $\beta_2(i,\xi) = \max_{k=0,\cdots,m-1}|\sigma_k^{(i)}(\xi)|$, and the following upper bound holds:

$$||\mathbf{H}_m^{-1}(\xi)||_{\infty} \leq m \cdot \max_{i=0,\cdots,m-1} \frac{(\delta_i(\xi)\prod_{\substack{j\neq i \\ 0\leq j\leq m-1}}(1+|\hat{a}(\frac{\xi+j}{m})|))^2}{|\hat{x}(\frac{\xi+i}{m})|}. \tag{3.14}$$

Proof Firstly, we prove the lower bound for $||\mathbf{H}_m^{-1}(\xi)||_{\infty}$. Denote the Vandermonde matrix $\mathbf{V}_m(\mathbf{w}(\xi))$ by $\mathbf{V}_m(\xi)$. Suppose $\mathbf{V}_m^{-1}(\xi) = (v_{ki})_{1\leq k,i\leq m}$ is the inverse of $\mathbf{V}_m(\xi)$, then by the inverse formula for a standard Vandermonde matrix,

$$v_{ki} = (-1)^{m-k}\sigma_{m-k}^{(i-1)}(\xi)\delta_{i-1}(\xi).$$

Let $\{e_i\}_{i=1}^m$ be the standard basis for $\mathbb{C}^m$ and $w_i(\xi) = \mathbf{V}_m^T(\xi)e_i$ for $i = 1, \cdots, m$. Since $|\hat{a}(\xi)| \leq 1$, we conclude that $||w_i||_{\infty} = 1$. Thus,

$$\begin{aligned}
||\mathbf{H}_m^{-1}(\xi)||_{\infty} &\geq \max_{i=1,\cdots,m}||\mathbf{H}_m^{-1}(\xi)w_i(\xi)||_{\infty} \\
&\geq m \cdot \max_{i=1,\cdots,m}\frac{||\mathbf{V}_m^{-1}(\xi)e_i||_{\infty}}{|\hat{x}(\frac{\xi+i}{m})|} \\
&= m \cdot \max_{i=0,\cdots,m-1}\frac{\beta_2(i,\xi)\delta_i(\xi)}{|\hat{x}(\frac{\xi+i}{m})|}.
\end{aligned} \tag{3.15}$$

On the other hand, using the factorization (2.4) and the upper bound norm estimate, we estimate for the inverse of a Vandermonde matrix in [46], we have shown that

$$\begin{aligned} ||\mathbf{H}_m^{-1}(\xi)||_\infty &\le m||\mathbf{V}_m^{-1}(\xi)||_\infty||((\mathbf{V}_m^{-1})^T(\xi))||_\infty||diag^{-1}(\mathbf{x}(\xi))||_\infty \\ &\le m \max_{i=0,\cdots,m-1} \frac{(\delta_i(\xi) \prod_{\substack{j\neq i \\ 0\le j\le m-1}} (1+|\hat{a}(\frac{\xi+j}{m})|))^2}{|\hat{x}(\frac{\xi+i}{m})|}. \end{aligned} \tag{3.16}$$

□

As an application of Proposition 3.3, the following theorem sheds some light on the dependence of $||\mathbf{H}_m^{-1}(\xi)||_\infty$ on m.

Theorem 3.2 *If $|\hat{x}(\xi)| \le M$ for every $\xi \in \mathbb{T}$, then $||\mathbf{H}_m^{-1}(\xi)||_\infty \ge O(2^m)$. Therefore, $||\mathbf{H}_m^{-1}(\xi)||_\infty \to \infty$ as $m \to \infty$.*

Proof We show this by proving $m \cdot \max_{i=0,\cdots,m-1} \delta_i(\xi) \ge O(2^m)$. Note $\beta_2(i,\xi) \ge |\sigma_0^{(i)}(\xi)| = 1$. By Eq. 3.15,

$$||\mathbf{H}_m^{-1}(\xi)||_\infty \ge m \cdot \frac{\max_{i=0,\cdots,m-1} \delta_i(\xi)}{M} = O(2^m), \tag{3.17}$$

the conclusion follows. Let $c(\xi) = \max_{i=0,\cdots,m-1} \delta_i(\xi)$ and note that

$$\begin{aligned} \frac{1}{c(\xi)^m} \le \prod_{i=0}^{m-1} \frac{1}{\delta_i(\xi)} &= \prod_{0\le i<j\le m-1} |\hat{a}(\tfrac{\xi+i}{m}) - \hat{a}(\tfrac{\xi+j}{m})|^2 \\ &= |det(\mathbf{V}_m(\xi)|^2. \end{aligned} \tag{3.18}$$

Since every entry of $\mathbf{w}(\xi)$ is contained in $[-1, 1]$, the Chebyshev points on $[-1, 1]$ maximize the determinant of Vandermonde matrix, see [24]. Therefore, by the formula for the determinant of a Vandermonde matrix on the Chebyshev points in [44], we get

$$|\det(\mathbf{V}_m(\xi))|^2 \le \frac{m^m}{2^{(m-1)^2}}.$$

By Eq. 3.18,

$$c(\xi) \ge \frac{2^{\frac{(m-1)^2}{m}}}{m}$$

which implies that $m \cdot c(\xi) \ge O(2^m)$. Hence by Eq. 3.17

$$||\mathbf{H}_m^{-1}(\xi)||_\infty \ge O(2^m) \to \infty, m \to \infty.$$

□

Remark 3.3 By our proof, we also see that $||\mathbf{H}_m^{-1}(\xi)||_\infty$ grows at least geometrically as m increases.

Summarizing, our results in this section suggest that

(1) For $0 \le k \le m-1$, the accuracy of recovering the node $\hat{a}(\frac{\xi+k}{m})$ not only depends on its separation from the other nodes $\delta_k(\xi)$ (see Definition 3.1), but also depends on the global minimal separation $\delta(\xi) = \max_{k=0,\cdots,m-1} \delta_k(\xi)$ among the nodes. Fix m and x, our estimates (3.12) and Eq. 3.16 suggest that the error $|\Delta_k(\xi)| = |\hat{\tilde{a}}(\frac{\xi+k}{m}) - \hat{a}(\frac{\xi+k}{m})|$ in the worst possible case could be proportional to $\delta_k(\xi)\delta^2(\xi)$. Our numerical experiment suggests this is sharp, see Fig. 2c and d.

(2) The accuracy of recovering all nodes is inversely proportional to the lowest magnitude of $\{\hat{x}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

(3) Increasing m may result in amplifying the error caused by the noise significantly, since the proof of Theorem 3.2 implies that $||\mathbf{H}_m^{-1}||_\infty$ grows at least geometrically when m increases. Thus, as m increases, our solutions likely become less robust to noise (see Fig. 2a and b).

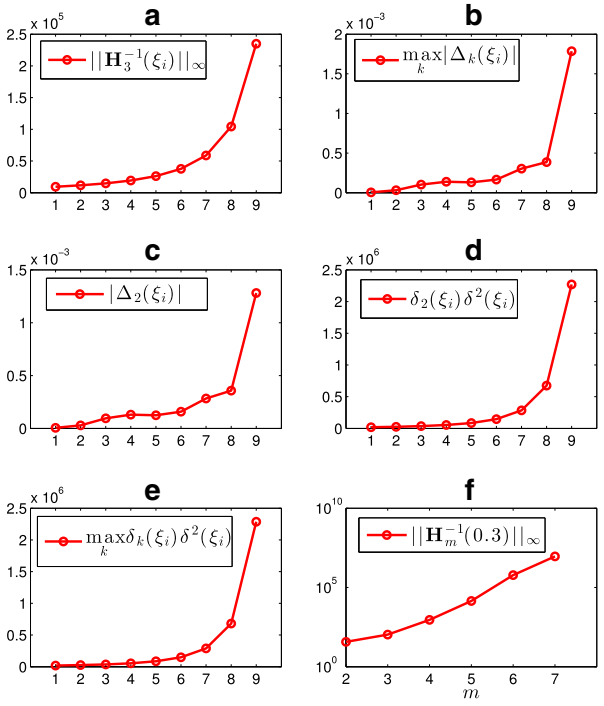

Fig. 2 Experiment Results

4 Numerical experiment

In this section, we provide some simple numerical simulations to verify the theoretical accuracy determined in Section 3.

4.1 Experiment setup

Suppose our filter a is around the center of radius 3. For example, let $a = (\cdots 0, 0.05, 0.4, 0.1, 0.4, 0.05, 0, \cdots)$ such that $\hat{a}(\xi) = 0.1 + 0.8\cos(2\pi\xi) + 0.1\cos(4\pi\xi)$, and $x = (\cdots, 0, 0.242, 0.383, 0.242, 0, \cdots)$ such that $\hat{x}(\xi) = 0.383 + 0.484\cos(2\pi\xi)$,and choose $m = 3$.

(1) In this experiment, we choose 9 points $[\xi_1, \cdots, \xi_9] = 0.49 : 0.001 : 0.498$ and calculate $\hat{y}_l(\xi_i)$ and the perturbed $\widetilde{\hat{y}}_l(\xi_i) = \hat{y}_l(\xi_i) + \epsilon_l$ for $l = 0, \cdots, 5$, where y_l is defined as in Eq. 1.1 and $\epsilon_l \sim 10^{-10}$.
(2) Use Algorithm 1 to calculate the roots of $q^{\xi}[z]$ and the perturbed roots of $\tilde{q}^{\xi}[z]$ respectively, then compute $|\Delta_k(\xi_i)| = |\tilde{\hat{a}}(\frac{\xi_i+k}{m}) - \hat{a}(\frac{\xi_i+k}{m})|$ for $k = 0, 1, 2$.
(3) Choosing $\xi = 0.3$, we compute $||\mathbf{H}_m^{-1}(0.3)||_\infty$ for $m = 2 : 1 : 7$.

4.2 Experiment results

In this subsection, we plot several figures to reflect the results of the experiment described above. The x-axis of Fig. 2a-e are set to be 1:9, which represent $\xi_1, \cdots, \xi_9$.

(1) **The dependence of $\max_k|\Delta_k(\xi)|$ on the infinity norm of $\mathbf{H_m^{-1}}(\xi)$.** Since the points $\xi_1, \cdots, \xi_9$ become closer to $\frac{1}{2}$, we expect the infinity norm of $\mathbf{H}_m^{-1}(\xi)$ to get larger and larger. Note that since m and x are fixed, the quantity $\mathbf{H}_m^{-1}(\xi)$ is the only significantly large item in the error estimations. We plot the value of $||\mathbf{H}_m^{-1}(\xi_i)||_\infty$ and $\max_k|\Delta_k(\xi_i)|$ for $i = 1, \cdots, 9$ in Fig. 2a and b. They exhibit almost the same behaviour and grow at a proportional rate. This indicates that the bigger $||\mathbf{H}_m^{-1}(\xi_i)||_\infty$ is, the bigger $\max_k|\Delta_k(\xi_i)|$ is.
(2) **Sharpness of Equations 3.5 and 3.14.** Our estimates (3.5) and Eq. 3.14 suggest that the error $|\Delta_k(\xi)|$ in the worst possible case could be proportional to $\delta_k(\xi)\delta^2(\xi)$. We plot the value of $|\Delta_2(\xi_i)|$ and $\delta_2(\xi)\delta^2(\xi)$ for $i = 1, \cdots, 9$ in Fig. 2c and d. It is indicated that $\Delta_2(\xi_i)$ grows approximately proportionally to the growth of $\delta_2(\xi_i)\delta^2(\xi_i)$, which suggests the sharpness of Eqs. 3.5 and 3.14. It is worthwhile to mention that the curve of $\max_k|\Delta_k(\xi_i)|$ coincides with the curve of $|\Delta_2(\xi_i)|$, and the curve of $\max_k \delta_k(\xi_i)\delta^2(\xi_i)$ coincides with the curve of $\delta_2(\xi_i)\delta^2(\xi_i)$. Since in this experiment, m and x are fixed, this also suggests that the quantity $\delta_k(\xi_i)\delta^2(\xi_i)$ essentially decides the accuracy. The bigger this quantity, the less accurate Algorithm 1 is.
(3) **The infinity norm of $\mathbf{H}_m^{-1}(\xi)$.** Recall in this experiment, we choose $m = 2, 3, \cdots, 6, 7$ and $\xi = 0.3$, and plot the value of $||\mathbf{H}_m^{-1}(0.3)||_\infty$ for different m.

The results are presented in Fig. 2f, in which the $y-$axis is set to be logarithmic. One sees that $||\mathbf{H}_m^{-1}(\xi)||_\infty$ grows geometrically.

5 Other numerical methods

In the following subsections, we will investigate the data structure of the Hankel matrix built from the spatiotemporal samples and present two algorithms based on the classical matrix pencil method and ESPRIT estimation method. These two classical methods are well known for their better numerical stability than the original Prony method.

5.1 Generalized Matrix Pencil Method

Let L and N be two integers satisfying $L \geq m$ and $N \geq L + m$. Then we define the $(N - L) \times 1$ column vector

$$\mathbf{h}_t(\xi) = [\hat{y}_t(\xi), \hat{y}_{t+1}(\xi), \cdots, \hat{y}_{N-L+t-1}(\xi)]^T,$$

and form the rectangular Hankel matrices

$$\mathbf{H}^{\xi}_{N-L,L+1} = \big[\mathbf{h}_0(\xi), \mathbf{h}_1(\xi), \cdots, \mathbf{h}_L(\xi)\big], \tag{5.1}$$

$$\mathbf{H}^{\xi}_{N-L,L}(s) = \mathbf{H}^{\xi}_{N-L,L+1}(1 : N - L, s + 1 : L + s), s = 0, 1.$$

Similar to the case $L = m$, for $s = 0, 1$,

$$\mathbf{H}^{\xi}_{N-L,L}(s) = V_{m,N-L}(\mathbf{w}(\xi))^T diag(\mathbf{x}(\xi)) diag(\mathbf{w}(\xi))^s V_{m,L}(\mathbf{w}(\xi)). \tag{5.2}$$

Recall that the superscripts "$*$" and "$+$" will denote the conjugate transpose and the pseudoinverse. The following lemma provides a foundation for the Generalized Matrix Pencil method.

Lemma 5.1 *Let N, L be two postive integers such that $m \leq L \leq N - m$. Assume $\xi \neq 0, \frac{1}{2}$ and $diag(\mathbf{x}(\xi))$ is invertible. The solutions to the generalized singular eigenvalue problem :*

$$(z\mathbf{H}^{\xi}_{N-L,L}(0) - \mathbf{H}^{\xi}_{N-L,L}(1))\mathbf{p}(\xi) = 0 \tag{5.3}$$

subject to $\mathbf{p}(\xi) \in \mathbf{R}(\mathbf{H}^{*\xi}_{N-L,L}(0))$, which denotes the column space of $\mathbf{H}^{*\xi}_{N-L,L}(0)$ are

$$z_i = \hat{a}(\frac{\xi + i - 1}{m})$$

$$\mathbf{p}(\xi) = \mathbf{p}_i(\xi) = \textit{ i-th column of } \mathbf{V}^{+}_{m,L}(\mathbf{w}(\xi))$$

for $i = 1, \cdots, m$.

Proof The proof follows from Eq. 5.2 via a similar method of proof to Theorem 2 in [28]. □

Proposition 5.1 *Let N, L be two postive integers such that $m \leq L \leq N-m$. Assume $\xi \neq 0, \frac{1}{2}$ and $diag(\mathbf{x}(\xi))$ is invertible. The $L \times L$ matrix $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$ has $\{\hat{a}(\frac{\xi+i}{m}), i = 0, \cdots, m-1\}$ and $L-m$ zeros as eigenvalues.*

Proof Left multiplying (5.3) by $\mathbf{H}^{+\xi}_{N-L,L}(0)$, we have

$$\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)\mathbf{p}_i(\xi) = z_i\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(0)\mathbf{p}_i(\xi). \tag{5.4}$$

By property of the pseudoinverse, $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(0)$ is the orthogonal projection onto $\mathbf{R}(\mathbf{H}^{*\xi}_{N-L,L}(0))$. Since $\mathbf{p}_i(\xi) \in \mathbf{R}(\mathbf{H}^{*\xi}_{N-L,L}(0))$, it is easy to see that the set $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ are m eigenvalues of $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$. Since the rank of $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$ is $m \leq L$, $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$ has $L-m$ zero eigenvalues. □

It is evident that one advantage of the matrix pencil method is the fact that there is no need to compute the coefficients of the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$. In this way, we need only solve a standard eigenvalue problem of a square matrix $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$. In order to compute $\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$. Inspired by the idea of Algorithm 5 in [29], we can employ the Singular Value Decomposition(SVD) based Matrix Pencil Method for Hankel matrices.

Lemma 5.2 *In addition to the conditions of Proposition 5.1, given the SVD of the Hankel matrix,*

$$\mathbf{H}^{\xi}_{N-L,L+1} = \mathbf{U}^{\xi}_{N-L}\Sigma^{\xi}_{N-L,L+1}\mathbf{W}^{\xi}_{L+1},$$

then

$$\mathbf{H}^{+\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1) = \mathbf{W}^{\xi\,+}_{L+1}(1:m, 1:L)\mathbf{W}^{\xi}_{L+1}(1:m, 2:L+1).$$

Proof This can be shown by direct computation and noticing that $\mathbf{H}^{\xi}_{N-L,L+1}$ has only m nonzero singular values. □

We summarize the Generalized Matrix Pencil Method in Algorithm 2. Note that the amount of computation required depends on the free parameter L. Numerical experiments show that the choice of L greatly affects the noise sensitivity of the eigenvalues. In terms of the noise sensitivity and computation cost, the best choice for L is between $\frac{N}{3}$ and $\frac{2N}{3}$ [29]. In our numerical examples, we choose L to be approximately $\frac{N}{3}$.

Algorithm 2 Generalized matrix pencil method (Based on SVD)

Input: $m \leq L \leq N-m$, $r \in \mathbb{N}$, $\{y_l\}_{l=0}^{N-1}$, $\xi(\neq 0, \frac{1}{2}) \in \mathbb{T}$.

1: Compute the Fourier transform of the measurement sequences $\{y_l\}_{l=0}^{N-1}$, build the Hankel matrix $\mathbf{H}^{\xi}_{N-L,L+1}$ and compute its SVD

$$\mathbf{H}^{\xi}_{N-L,L+1} = \mathbf{U}^{\xi}_{N-L} \Sigma^{\xi}_{N-L,L+1} \mathbf{W}^{\xi}_{L+1}.$$

2: Compute the eigenvalues of $W^{\xi}{}^{+}_{L+1}(1:m, 1:L) W^{\xi}_{L+1}(1:m, 2:L+1)$.

3: Delete $L-m$ smallest values in modulus (zeros in noise-free case) from the eigenvalues. Order the remaining eigenvalues by the monotonicity and symmetry condition of to get $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

Output: $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

5.2 Generalized ESPRIT Method

The original ESPRIT Method relies on a particular property of Vandermonde matrices known as rotational invariance [22]. By the factorization (5.2), we have seen that the Hankel data matrix $\mathbf{H}^{\xi}_{N-L,L+1}$ containing successive spatiotemporal data of the evolving states is rank deficient and that its range space, known as the signal subspace, is spanned by the Vandermonde matrix generated by $\{\hat{a}(\frac{\xi+i}{m}), i = 0, \cdots, m-1\}$. Hence, we can generalize the ESPRIT algorithm based on SVD for estimating $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ in our setting. We summarize this method in Algorithm 3

Algorithm 3 Generalized ESPRIT algorithm

Input: $m \leq L \leq N-m$, $r \in \mathbb{N}$, $\{y_l\}_{l=0}^{N-1}$, $\xi(\neq 0, \frac{1}{2}) \in \mathbb{T}$.

1: Compute the Fourier transform of the measurement sequences $\{y_l\}_{l=0}^{N-1}$ and form the Hankel matrix $\mathbf{H}^{\xi}_{N-L,L+1}$.

2: Compute the SVD of $\mathbf{H}^{\xi}_{N-L,L+1} = \mathbf{U}^{\xi}_{N-L} \Sigma^{\xi}_{N-L,L+1} \mathbf{W}^{\xi}_{L+1}$.

3: Compute the $m \times m$ spectral matrix $\Phi(\xi)$ by solving the linear system

$$\mathbf{U}^{\xi}_{N-L}(1:N-L-1, 1:m)\Phi(\xi) = \mathbf{U}^{\xi}_{N-L}(2:N-L, 1:m)$$

and estimate the eigenvalues of $\Phi(\xi)$.

4: Order the eigenvalues by the monotonicity and symmetry condition of $\hat{a}$ to get $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

Output: $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

5.3 Data preprocessing using Cadzow Denoising Method

It has been shown in the previous sections that the Hankel matrix $\mathbf{H}^{\xi}_{N-L,L+1}(m \leq L \leq N-m)$ has two key properties in the noise-free case under appropriate hypotheses:

(1) It has rank m.
(2) It is Toeplitz.

In the noisy case, these two properties are not initially satisfied simultaneously. $\mathbf{H}^{\xi}_{N-L,L+1}$ is very sensitive to noise: numerical experiments show that even very small noise ($\sim 10^{-10}$) will change its rank dramatically. To further improve robustness, we use an iterative method devised by Cadzow [27] to preprocess the noisy data which guarantees a Hankel matrix with the above key properties.

Algorithm 4 Cadzow iterative denoising method

Input: $m \leq L \leq N-m$, $\{\hat{\tilde{y}}_l(\xi)\}_{l=0}^{N-1}$, $\xi(\neq 0, \frac{1}{2}) \in \mathbb{T}$.

1: Build the Hankel matrix $\mathbf{H}^{\xi}_{N-L,L+1}$ from $\{\hat{\tilde{y}}_l(\xi)\}_{l=0}^{N-1}$ and perform the SVD. Let $\lambda_1, \cdots, \lambda_K$ be its singular values, $K = \min\{N-L, L+1\}$.
2: Set ϵ to be a small positive number.
3: **while** $\frac{\lambda_{m+1}}{\lambda_m} \geq \epsilon$ **do**
4: Enforce the rank m of $\mathbf{H}^{\xi}_{N-L,L+1}$ by setting the $K-m$ smallest singular values to zero.
5: Enforce the Toeplitz structure on $\mathbf{H}^{\xi}_{N-L,L+1}$ by averaging the entries along the diagonals.
6: **end while**
7: Extract the denoised Fourier data $\{\hat{\tilde{y}}_l(\xi)\}_{l=0}^{N-1}$ from the first column and the last row of $\mathbf{H}^{\xi}_{N-L,L+1}$

Output: Denoised Fourier data $\{\hat{\tilde{y}}_l(\xi)\}_{l=0}^{N-1}$ and Hankel matrix $\mathbf{H}^{\xi}_{N-L,L+1}$.

The procedure of Algorithm 4 is guaranteed to converge to a matrix which exhibits the desired two key properties [27]. The iterations stop whenever the ratio of the $(m+1)$-th singular value to the m-th one, falls below a predetermined threshold. Since Algorithm 1 does not perform well for large noise, we can combine the Algorithm 1 and Algorithm 4 to recover the Fourier spectrum of a and improve the performance.

6 Numerical examples

In this section, we present a numerical example to illustrate the effectiveness and robustness of the proposed Algorithms.

Example 6.1 Let the filter

$$a = (\cdots, 0, 0.25, 0.5, 0.25, 0, \cdots)$$

so that $\hat{a}(\xi) = 0.5 + 0.5\cos(2\pi\xi)$. $\hat{a}$ is approximately Gaussian on $[-\frac{1}{2}, \frac{1}{2}]$. Let the initial signal x be a conjugate symmetric vector given by $x(0) = 0.75, x(1) = \bar{x}(-1) = 0.8976 + 0.4305i$,and $x(2) = \bar{x}(-2) = 0.9856 - 0.1682i$ so that $\hat{x}(\xi) = 0.75+2Re((0.9856-0.1682i)e^{-4\pi i\xi}+(0.8976-0.4305i)e^{-2\pi i\xi})$. The subsampling factor m is set to be 5. Given the Fourier data of the spatiotemporal samples $\{\hat{y}_l\}_{l=0}^{N-1}$, we add independent uniform distributed noise $\epsilon_l \sim U(-\epsilon, \epsilon)$ to the Fourier data $\hat{y}_l$

Table 1 Numerical results

Algorithm	m	N	L	e_{best}	e_{worst}	MSE
1	5	10	5	0.4e-07	0.19e-05	0.14e-05
	5	15	5	0.27e-08	0.24e-06	0.17e-06
	5	20	5	0.85e-09	0.18e-06	0.13e-06
	5	25	5	0.46e-09	0.17e-06	0.12e-06
	5	50	5	0.19e-09	0.15e-06	0.10e-06
	5	100	5	0.17e-09	0.14e-06	0.96e-07
	5	200	5	0.17e-09	0.14e-06	0.97e-07
1+4	5	10	5	0.38e-07	0.18e-05	0.13e-05
	5	15	5	0.13e-08	0.13e-06	0.09e-06
	5	20	5	0.15e-09	0.58e-07	0.41e-07
	5	25	5	0.49e-10	0.42e-07	0.29e-07
	5	50	5	0.84e-11	0.29e-07	0.20e-07
	5	100	5	0.77e-11	0.30e-07	0.21e-07
	5	200	5	0.97e-11	0.30e-07	0.21e-07
2	5	15	5	0.17e-08	0.16e-06	012e-06
	5	20	6	0.25e-09	0.74e-07	0.53e-07
	5	25	8	0.69e-10	0.47e-07	0.33e-07
	5	50	16	0.64e-11	0.27e-07	0.19e-07
	5	100	33	0.28e-11	0.24e-07	0.17e-07
	5	200	66	0.26e-11	0.24e-07	0.17e-07
3	5	15	5	0.17e-08	0.16e-06	0.11e-06
	5	20	6	0.21e-09	0.66e-07	0.46e-07
	5	25	8	0.62e-10	0.45e-07	0.32e-07
	5	50	16	0.60e-11	0.27e-07	0.19e-07
	5	100	33	0.28e-11	0.24e-07	0.17e-07
	5	200	66	0.26e-11	0.24e-07	0.17e-07

for $l = 0, \cdots, N-1$. Recalling that $|\Delta_k(\xi)| = |\tilde{\hat{a}}(\frac{\xi+k}{m}) - \hat{a}(\frac{\xi+k}{m})|$, we define the relative error

$$\mathbf{e}_k(\xi) = \frac{|\Delta_k(\xi)|}{\max\limits_k |\hat{a}(\frac{\xi+k}{m})|}$$

for $k = 0, 1, \cdots, m-1$. The best case error is set to be $\mathbf{e}_{best}(\xi) = \min\limits_k \mathbf{e}_k(\xi)$ and the worst case error is set to be $\mathbf{e}_{worst}(\xi) = \max\limits_k \mathbf{e}_k(\xi)$. Additionally, we define the mean square error via

$$\mathbf{MSE}^2(\xi) = \frac{\sum\limits_{k=0}^{m-1} |\Delta_k(\xi)|^2}{\sum\limits_{k=0}^{m-1} |\hat{a}(\frac{\xi+k}{m})|^2}.$$

Then we apply our Algorithm 1, Algorithms 1+Algorithm 4, Algorithm 2 and Algorithm 3 to the case when $\xi = 0.4$. For several parameters N and L, the resulting errors (average over 100 experiments) are presented in Table 1. As the bound ϵ in the algorithms we use 10^{-10}. It is shown in the Table 1 that increasing the temporal samples, i.e. N, will help reduce the errors. The new proposed algorithms have better performance than Algorithm 1 if given more spatiotemporal data.

Table 2 Numerical results for a_1

Algorithm	m	N	L	e_{best}	e_{worst}	MSE
1	5	15	5	0.24e-08	0.21e-04	0.15e-04
	5	50	5	0.33e-09	0.13e-04	0.90e-05
	5	100	5	0.34e-09	0.13e-04	0.93e-05
	5	200	5	0.32e-09	0.12e-04	0.86e-05
1+4	5	15	5	0.93e-09	0.94e-05	0.67e-05
	5	50	5	0.16e-10	0.25e-05	0.18e-05
	5	100	5	0.16e-10	0.26e-05	0.18e-05
	5	200	5	0.15e-10	0.27e-05	0.19e-05
2	5	15	5	0.14e-08	0.13e-04	0.91e-05
	5	50	16	0.12e-10	0.22e-05	0.16e-05
	5	100	33	0.84e-11	0.22e-05	0.16e-05
	5	200	66	0.81e-11	0.23e-05	0.16e-05
3	5	15	5	0.13e-08	0.11e-04	0.81e-05
	5	50	16	0.11e-10	0.21e-05	0.15e-05
	5	100	33	0.83e-11	0.22e-05	0.16e-05
	5	200	66	0.81e-11	0.23e-05	0.16e-05

Example 6.2 In this example, we explore the relationship between the bandwidth of the filter a and the performance of the proposed algorithms. Let

$$a_1 = (\cdots, 0, 0.05, 0.2, 0.5, 0.2, 0.05, 0, \cdots)$$

and

$$a_2 = (\cdots, 0, 0.001, 0.002, 0.05, 0.06, 0.24, 0.3, 0.24, 0.06, 0.05, 0.002, 0.001, 0, \cdots),$$

and we keep other experiment settings being the same with those in Example 6.1. The bandwidth of a_1 is 3 and the bandwidth of a_2 is 6. $\hat{a}_1(\xi) = 0.5 + 0.4\cos(2\pi\xi) + 0.1\cos(4\pi\xi)$ and $\hat{a}_2(\xi) = 0.3 + 0.48\cos(2\pi\xi) + 0.12\cos(4\pi\xi) + 0.1\cos(6\pi\xi) + 0.004\cos(8\pi\xi) + 0.002\cos(10\pi\xi)$. Recall that the quantities $\{\delta_k(\xi) : k = 0, 1, \cdots, 4\}$ defined in Definition 3.1 play an essential role in perturbation analysis of Prony type methods studied in Section 3. Define $\delta(\xi) = \max_k \delta_k(\xi)$, we compute quantities $\delta_k(\xi)\delta^2(\xi)$ for a_1 and a_2 at $\xi = 0.4$. We find that the quantity $\delta_k(\xi)\delta^2(\xi)$ associated with a_1 is almost 10 times the quantity $\delta_k(\xi)\delta^2(\xi)$ associated with a_2 for $k = 0, 1, \cdots, 4$. Despite the fact that a_1 has narrower band than a_2, we expect that the error of recovering $\{\hat{a}_1(\frac{\xi+k}{m}) : k = 0, 1, \cdots, m-1, \xi = 0.4\}$ is larger than the error of recovering $\{\hat{a}_2(\frac{\xi+k}{m}) : k = 0, 1, \cdots, m-1, \xi = 0.4\}$. We perform experiments, we summarize the resulting errors (average over 100 experiments) in Table 2 for a_1 and Table 3 for a_2, in which our expectation is fulfilled. From this numerical example, we see that, it is the separation among the recovering nodes that essentially determines the accuracy. It would be interesting to explore how the bandwidth interplay with the separations among roots.

Table 3 Numerical Results for a_2

Algorithm	m	N	L	e_{best}	e_{worst}	MSE
1	5	15	5	0.49e-09	0.16e-05	0.13e-05
	5	50	5	0.48e-09	0.16e-05	0.13e-05
	5	100	5	0.40e-09	0.14e-05	0.12e-05
	5	200	5	0.43e-09	0.16e-05	0.13e-05
1+4	5	15	5	0.30e-09	0.11e-05	0.92e-06
	5	50	5	0.21e-09	0.10e-05	0.86e-06
	5	100	5	0.20e-09	0.10e-05	0.84e-06
	5	200	5	0.21e-09	0.98e-06	0.82e-06
2	5	15	5	0.33e-09	0.12e-05	0.10e-05
	5	50	16	0.21e-09	0.10e-05	0.86e-06
	5	100	33	0.20e-09	0.10e-05	0.84e-06
	5	200	66	0.21e-09	0.98e-06	0.82e-06
3	5	15	5	0.32e-09	0.11e-05	0.96e-06
	5	50	16	0.21e-09	0.10e-05	0.86e-06
	5	100	33	0.20e–09	0.10e-05	0.84e-06
	5	200	66	0.21e-09	0.98e-06	0.82e-06

7 Conclusion

In this paper, we investigate the conditions under which we can recover a typical low pass convolution filter $a \in \ell^1(\mathbb{Z})$ and a vector $x \in \ell^2(\mathbb{Z})$ from the combined regular subsampled version of the vector $x, \cdots, A^{N-1}x$ defined in Eq. 1.1, where $Ax = a * x$. We show that if one doubles the amount of temporal samples needed in [11] to recover the signal propagated by a known filter, one can almost surely solve the problem even if the filter is unknown. We first propose an algorithm based on the classical Prony method to recover a finite impulse response filter and signal in the case where an upper bound for their support is known In particular, we have done a first order perturbation analysis and the estimates are formulated in simple geometric terms involving the Fourier spectral function of a, x and m, shedding some light on the structure of the problem. We obtain a lower bound estimation for infinity norm of $\mathbf{H}_m^{-1}(\xi)$ in terms of m. Then we propose several other algorithms, which can make use of more temporal samples and increase robustness of the recovery in the presence of noise. Potential applications include One-Chip Sensing (in which sensors are placed inside chips to measure voltage, current, and temperature e.g. to avoid overheating) and source localization where there is an evolving state (e.g. sound acquisition via microphones).

Acknowledgments The author would like to thank Akram Aldroubi for his endless encouragement and very insightful comments in the process of creating this work. The author is also indebted to Ilya Kristhal and Yang Wang for their helpful discussions related to this work. The author would like to thank anonymous reviewers for their valuable advices to improve the quality of this manuscript. In addition, the author thanks Keaton Hamm, Armenak Petrosyan and Sayan Das for their useful suggestions to improve the presentation of this manuscript.

References

1. Aldroubi, A., Krishtal, I.: Krylov subspace methods in dynamical sampling. arXiv:1412.1538 (2014)
2. Aldroubi, A., Molter, U., Cabrelli, C., Tang, S.: Dynamical Sampling. arXiv:1409.8333. To appear in Applied And Computational Harmonic Analysis
3. Adcock, B., Hansen, A.: A generalized sampling theorem for stable reconstructions in arbitrary bases. J. Fourier Anal. Appl. **18**(4), 685–716 (2012)
4. Sun, Q.: Nonuniform average sampling and reconstruction of signals with finite rate of innovation. SIAM J. Math. Anal. **38**(5), 1389–1422 (2006/07). electronic
5. Adcock, B., Hansen, A.: Generalized sampling and the stable and accurate reconstruction of piecewise analytic functions from their Fourier coefficients. Math. Comp. **84**, 237–270 (2015)
6. Slepian, D., Wolf, J.: Noiseless coding of correlated information sources. IEEE Trans. Inf. Theory **19**, 471 –480 (1973)
7. Gilbert, A., Indyk, P., Iwen, M.A., Schmidt, L.: Recent Developments in the Sparse Fourier Transform. IEEE Signal Process. Mag. **31**(5), 91–100 (2014)
8. Iwen, M.A.: Combinatorial sublinear-time Fourier algorithms. Found. Comput. Math. **10**, 303–338 (2010)
9. Iwen, M.A.: Improved approximation guarantees for sublinear-time Fourier algorithms. Appl. Comput. Harmon. Anal. **34**, 57–82 (2013)
10. Lawlor, D., Wang, Y., Christlieb, A.: Adaptive sub-linear time Fourier algorithms. Adv. Adapt. Data Anal. **5**(1) (2013)
11. Aldroubi, A., Davis, J., Krishtal, I.: Exact reconstruction of signals in evolutionary systems via spatiotemporal trade-off. J. Fourier Anal. Appl. **21**(1), 11–31 (2015)

12. Peter, T., Plonka, G.: A generalized Prony method for reconstruction of sparse sums of eigenfunctions of linear operators. Inverse Probl. **29**, 025001 (2013)
13. Sun, Q.: Frames in spaces with finite rate of innovation. Adv. Comput. Math. **28**(4), 301–329 (2008)
14. Aldroubi, A., Krishtal, I., Weber, E. In: Balan, R., Begue, M., Benedetto, J., Czaja, W., Okodujou, K. (eds.): Finite dimensional dynamical sampling: an overview. Excursions in harmonic analysis, vol. 3. Applied Numerical Harmonic Analysis, Birkhäuser/Springer, New York (2015). To appear
15. Aldroubi, A., Gröchenig, K.: Nonuniform sampling and reconstruction in shift-invariant spaces. SIAM Rev. **43**, 585–620 (2001)
16. Bass, R.F., Gröchenig, K.: Relevant sampling of bandlimited functions, Illinois J. Math To appear (2012)
17. Christensen, J., Ólafsson, G.: Sampling in spaces of bandlimited functions on commutative spaces. In: Applied and Numerical Harmonic Analysis book Excursions in Harmonic Analysis, vol. 1, pp. 35–69 (2013)
18. Jorgensen, P., Tian, F.eng.: Discrete reproducing kernel Hilbert spaces: Sampling and distribution of Dirac-masses. arXiv:1501.02310
19. Iosevich, A., Mayeli, A.: Exponential bases, Paley-Wiener spaces, and applications. J. Funct. Anal. **268**(2), 363–375 (2015)
20. Adcock, B., Hansen, A., Herrholz, E., Teschke, G.: Generalized sampling: extensions to frames and inverse and ill-posed problems. Inverse Probl. **29**, 015008
21. Aldroubi, A., Davis, J., Krishtal, I.: Dynamical sampling: time space trade-off. Appl. Comput. Harmon. Anal. **34**(3), 495–503 (2013)
22. Roy, R., Paulraj, A., Kailath, T.: ESPRIT-A subspace rotation ap- proach to estimation of parameters of cisoids in noise. IEEE Trans. Acoust., Speech Signal Process. **34**(5), 1340–1342 (1986)
23. Aceska, R., Tang, S.: Dynamical sampling in hybrid shift invariant spaces. In: Furst, V., Kornelsen, K., Weber, E. (eds.) Operator Methods in Wavelets, Tilings, and Frame, vol. 626. Contemporary Mathematics, American Mathematics Society, Providence (2014). To appear
24. Borodchov, S., Hardin, D., Saff, E.B.: Minimal Discrete Energy on the Sphere and other Manifolds, Springer. To appear
25. Aceska, R., Aldroubi, A., Davis, J., Petrosyan, A.: Dynamical sampling in shift invariant spaces. In: Mayeli, A., Iosevich, A., Jorgensen, P., Ólafsson, G. (eds.) Commutative and Noncommutative Harmonic Analysis and Applications, vol. 603, pp. 139–148. Contemporary Mathematics, American Mathematical Society, Providence (2013)
26. Reise, G., Matz, G., Gröchenig, K.: Distributed field reconstruction in wireless sensor networks based on hybrid shift-invariant spaces. IEEE Trans Signal Processi **60**(10), 5426–5439 (2012)
27. Cadzow, J.A.: Signal enhancement—A composite property mapping algorithm. IEEE Trans. Acoust., Speech Signal Process. **36**, 49–67 (1988)
28. Hua, Y., Sarar, T.K.: Matrix Pencil Method for Estimating Parameters of Exponentially Damped/ Undamped Sinusoids in Noise. IEEE Trans. Acoust., Speech Signal Process. **38**, 814–824 (1990)
29. Hua, Y.ingbo., Sarar, T.K.: On SVD for Estimating Generalized Eigenvalues of Singular Matrix Pencil in Noise. IEEE Trans. Signal Process. **39**(4), 892–900 (1991)
30. Reise, G., Matz, G.: Distributed sampling and reconstruction of non-bandlimited fields in sensor networks based on shift-invariant spaces. In: Proceedings ICASSP, pp. 2061–2064, Taipeh, Taiwan (2009)
31. Reise, G., Matz, G.: Reconstruction of time-varying fields in wireless sensor networks using shift-invariant spaces: interative algorithms and impact of sensor localization errors. In: Proceedings SPAWC, pp. 1–5, Marrakech, Morocco (2010)
32. Jorgensen, P.: A sampling theory for infinite weighted graphs. Opuscula Math. **31**(2), 209–236 (2011)
33. Nashed, M.Z., Sun, Q.: Sampling and reconstruction of signals in a reproducing kernel subspace of $L_p(R^d)$. J. Funct Anal. **258**(7), 2422–2452 (2010)
34. Lu, Y., Vetterli, M.: Spatial super-resolution of a diffusion field by temporal oversampling in sensor networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009, pp. 2249 –2252 (2009)
35. Hormati, A., Roy, O., Lu, Y., Vetterli, M.: Distributed sampling of signals linked by sparse filtering: theory and applications. IEEE Trans. Signal Process. **58**(3), 1095–1109 (2010)

36. Daubechies, I.: Ten lectures on wavelets CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1992)
37. Han, D., Nashed, M.Z., Sun, Q.: Sampling expansions in reproducing kernel Hilbert and Banach spaces. Numer. Funct. Anal. Optim. **30**(9-10), 971–987 (2009)
38. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: A survey. Comput. Netw. **38**, 393–422 (2002)
39. Blu, T., Dragotti, P.L., Vetterli, M., Marziliano, P., Coulot, L.: Sparse sampling of signal innovations. IEEE Signal Process. Mag. **25**, 31–40 (2008)
40. Badeau, R., David, B., Richard, G.: High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials. IEEE Trans. Signal Process. **54**(4), 1341–1350 (2006)
41. Vinberg, E.B.: A course in algebra. American Mathematical Society, Providence, R.I (2003). ISBN 0-8218-3413-4
42. Benedetto, J.J., Ferreira, P.J.S.G. (eds.): Modern sampling theory. Modern sampling theory, Applied and Numerical Harmonic Analysis, Birkhäuser Boston Inc., Boston, MA (2001)
43. Nashed, M.Z.: Inverse problems, moment problems, signal processing un menage a trois, Mathematics in science and technology, pp. 2–19. World Scientific Publications, Hackensack, NJ (2011)
44. Eisinberg, A., Franzé, G., Pugliese, P.: Vandermonde matrices on Chebyshev points. Linear Algebra Appl. **283**(1–3), 205–219 (1998)
45. Timan, A.: Theory of approximation of functions of a real variable. Courier Dover Publications, 37 (1963)
46. Gautschi, W.: Norm estimates for inverses of Vandermonde matrices. Numer. Math. **23**, 337–347 (1975)
47. Gautschi, W.: How (Un)stable Are Vandermonde Systems. Asymptotic and computational analysis (Winnipeg, MB, 1989) Lecture Notes in Pure and Applications of Mathematics, vol. 124, pp. 193–210. Dekker, New York (1990)