Dynamical Sampling

By

Sui Tang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematics

August, 2016

Nashville, Tennessee

Approved:

Akram Aldroubi, Ph.D

Douglas P. Hardin, Ph.D

Akos Ledeczi, Ph.D

Alexander M. Powell, Ph.D

Edward B. Saff, Ph.D

To my dear parents,

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor Prof. Akram Aldroubi. I feel that I am the most fortunate graduate student to work under his advisorship.In the last five years, he has taught me how to ask questions, make baby steps and pursue my own ideas persistently. The joy and enthusiasm he has for his research was contagious for me, and helped me overcome crisis situations during tough times in the Ph.D. pursuit. He is also very supportive for my academic career. He has given me a lot of opportunities for meeting people and presenting my work in conferences, and made me more visible in the community. Besides mathematics, I have enjoyed a lot of fun playing Ping-Pong with him every time after our discussions, and have learned swimming under his encouragement. I appreciate all his contributions of precious time, insightful ideas, and funding to make my Ph.D. experience productive and stimulating.

I would like to thank all faculty members in my committee for their support on crucial times. I would like to sincerely thank Prof. Alex Powell, Prof. Ed Saff, Prof. Doug Hardin and Prof. Larry Schumaker for the inspiring conversations and interesting mathematics that they have offered. Additionally, I would like to thank Prof. Carlos Cabrelli, Prof. Ursula Molter, Prof. Ilya Krishtal, Prof. Yang Wang, Zhengwei Liu and Oleksandr Vlasiuk. I benefit a lot from insightful discussions with them. I am grateful to Prof. Yang Wang for providing me with opportunities to visit mathematics department of MSU and HKUST. I am also grateful to Prof. Mauro Maggioni for providing me with opportunities to visit Duke University.

I am grateful to all members in HAG research group for the heuristic discussions we have had and happy get together time we have spent. Additionally, I would like to thank my academic sibling, Xuemei Chen. She has been very supportive. She helped me a lot in settling down when I first came to Vanderbilt and offered me a lot of invaluable suggestions on my graduate study. I am very grateful for her time and effort in reading my paper and

giving me suggestions. I also want to sincerely thank Wenjing Liao and Haichao Wang for their invaluable suggestions and support for my academic career.

I am thankful for all my friends at Vanderbilt University, for making my graduate life a wonderful and enjoyable period in my life. I would like to give special thanks to Xi Huo, who has given me a lot of help in daily life and introduced me a lot of interesting places in Nashville.

Most importantly, none of this work would have been accomplished without the love and support of my parents. Eternal gratitude goes to Yunxiang Ren for his faithful support and dedication on this mathematical adventure.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 Motivation

In Wireless Sensor Networks (WSN), a large amount of sensors are distributed to gather information about a field to be monitored, such as pollution, temperature or pressure. The goal is to exploit the evolutionary structure and the placement of sensors to reconstruct an unknown field. In practice, increasing the spatial sampling density is usually much more expensive than increasing the temporal sampling density, and it is not always possible to place sampling devices at all desired locations. These issues motivate us to propose a spatiotemporal sampling framework to do sampling and reconstruction in evolution processes. The idea is exploiting spatiotemporal correlation and using a reduced number of sensors with each being activated more frequently. In other words, we seek to compensate the insufficient spatial sampling rate by oversampling in time and achieve spatiotemporal trade off. Lu, Vetterli, and their collaborators study the spatiotemporal trade off in heat diffusion processes (e.g. [35]). Later, Aldroubi and his collaborators develop a mathematical framework of dynamical sampling to study the spatiotemporal trade off in case of regular subsampling in the discrete spatially invariant evolution systems([5, 6]). In practice, there are many other types of evolution systems and it is often the case that the sensors are scattered irregularly, such as dropped from an airplane. So it is important to extend the dynamical sampling to more general evolution systems and consider the irregular subsampling and possibly random subsampling.

1.2 General Problem Formulation

We mainly consider the dynamical sampling problem in the separable Hilbert space. The general dynamical sampling problem can be stated as follows: Let f be a function

in a separable Hilbert space $\mathscr{H}$, e.g., $\mathbb{C}^d$ or $\ell^2(\mathbb{N})$, and assume that f evolves through an evolution operator $A : \mathscr{H} \to \mathscr{H}$ so that the function at time n has evolved to become $f^{(n)} = A^n f$. We identify $\mathscr{H}$ with $\ell^2(I)$ where $I = \{1, \dots, d\}$ in the finite dimensional case, and $I = \mathbb{N}$(or $\mathbb{Z}$) in the infinite dimensional case. We denote by $\{e_i\}_{i \in I}$ the standard basis of $\ell^2(I)$.

The time-space sample at time $t \in \mathbb{N}$ and location $p \in I$, is the value $A^t f(p)$. In this way we associate to each pair $(p,t) \in I \times \mathbb{N}$ a sample value.

At time $t = n$, we sample f at the locations $\Omega_n \subseteq I$ resulting in the measurements $\{f^{(n)}(i) : i \in \Omega_n\}$. Here $f^{(n)}(i) = < A^n f, e_i >$.

In general, the measurements $\{f^{(0)}(i) : i \in \Omega_0\}$ that we have from our original signal $f = f^{(0)}$ will contain in general insufficient information to recover f. In other words, f is undersampled. So we will need some extra information from the iterations of f by the operator A: $\{f^{(n)}(i) = A^n f(i) : i \in \Omega_n\}$. Again, for each n, the measurements $\{f^{(n)}(i) : i \in \Omega_n\}$ that we have by sampling our signals $A^n f$ at Ω_n are insufficient to recover $A^n f$ in general.

Several questions arise. Will the combined measurements $\{f^{(n)}(i) : i \in \Omega_n\}$ contain in general all the information needed to recover f (and hence $A^n f$)? How many iterations L will we need (i.e., $n = 1, \dots, L$) to recover the original signal? What are the right "spatial" sampling sets Ω_n we need to choose in order to recover f? In what way all these questions depend on the operator A? The general dynamical sampling problem can then be described as:

Problem 1.2.1 (Spatiotemporal Trade off)**.** Under what conditions on the operator A, and a set $S \subseteq I \times \mathbb{N}$, can every f in the Hilbert space H be recovered in a stable way from the samples in S.

The name of the above Problem 1.2.1 comes from the fact that in many cases it is possible to provide the same information about the initial state from a reduced number of devices activated more frequently. Another important problem arises when the evolution

operators are themselves unknown (or partially known).

Problem 1.2.2 (System Identification in Dynamical Sampling)**.** Assume A is unknown or parially known, under what conditions on the operator A, and a set $S \subseteq I \times \mathbb{N}$, can the key parameters of A and even every f in the Hilbert space H be recovered in a stable way from the samples in S.

1.3 Connection to other fileds

The dynamical sampling problem has similarities to other areas of mathematics. For example, in wavelet theory [10, 17, 18, 27, 36, 39, 43], a high-pass convolution operator H and a low-pass convolution operator L are applied to the function f. The goal is to design operators H and L so that reconstruction of f from samples of Hf and Lf is feasible. In dynamical sampling there is only one operator A, and it is applied iteratively to the function f. Furthermore, the operator A may be high-pass, low-pass, or neither and is given in the problem formulation, not designed.

In inverse problems (see [38] and the references therein), a single operator B, that often represents a physical process, is to be inverted. The goal is to recover a function f from the observation Bf. If B is not bounded below, the problem is considered an ill-posed inverse problem. Dynamical sampling is different because $A^n f$ is not necessarily known for any n; instead f is to be recovered from partial knowledge of $A^n f$ for many values of n. In fact, the dynamical sampling problem can be phrased as an inverse problem when the operator B is the operation of applying the operators $A, A^2, \ldots, A^L$ and then subsampling each of these signals accordingly on some sets Ω_n for times $t = n$.

The methods that we develop for studying the spatiotemporal trade off Problem 1.2.1 are related to methods in spectral theory, operator algebras, and frame theory [3, 12, 14, 16, 20, 23, 24, 25, 44]. For example, the proof of Theorem 2.4.15, below, uses the newly proved Kadison-Singer/Feichtinger conjecture [37]. Another example is the existence of cyclic vectors that form frames, which is related to Carleson's Theorem for interpolating

sequences in the Hardy space $H^2(\mathbb{D})$ (c.f., Theorem 2.4.16). Various versions of Problem 1.2.2 of Dynamical sampling exhibit features similar to many fundamental problems in other area such as super resolution, blind deblurring. But even in the most basic case, they necessitate new theoretical and algorithmic techniques.

1.4 Overview and Organization

In chapter 2, we consider the spatiotemporal trade off problem in both finite dimensional and infinite dimensional separable Hilbert spaces. We completely solve Problem 1.2.1 in the finite dimensional spaces, and for a large class of self adjoint operators in infinite dimensional spaces. We give a characterization specifying what the right spatial sampling sets Ω we need to choose are, how many iterations l_i we need for each $i \in \Omega$, and in what way they depend on the operator A to recover the original signal f. The work in this section is joint work with Akram Aldroubi, Carlos Cabrelli and Ursula Molter, and appears in [4].

In chapter 3, we consider the system identification problem of dynamical sampling in the infinite dimensional spatially invariant evolution processes. We consider a regular spatiotemporal subsampling scheme and show that if the amount of temporal samples is equal to the double of amount of samples for the case when the convolution operator A is known, one can almost surely solve Problem 1.2.2 for this specal case. We propose several algorthms for the case when both the convolution kernel of A and initial signal are compactly supported. The work in this section is independent work and can be found in [45].

In chapter 4, we look back on the spatiotemporal trade off problem in the spatially invariant systems and extend one variable results in case of regular subsampling in [5, 6] to the multivariable setting. This work is motivated by the fact that in industrial applications, the observed time variant signals are described by at least two variables. The work in this section is joint work with Roza Aceska and Armenak Petrosyan, and appears in [2].

Chapter 2

SPATIOTEMPORAL TRADE OFF PROBLEM IN HILBERT SPACE

2.1 Problem Formulation

In this section, we formulate a special case of spatiotemporal trade off Problem 1.2.1 and the goal of this chapter is to understand completely this problem that we can formulate as:

Let A be the evolution operator acting in $\ell^2(I)$, $\Omega \subseteq I$ a fixed set of locations, and $\{l_i : i \in \Omega\}$ where l_i is a positive integer or $+\infty$.

Problem 2.1.1. Find conditions on A, Ω and $\{l_i : i \in \Omega\}$ such that any vector $f \in \ell^2(I)$ can be recovered from the samples $Y = \{f(i), Af(i), \ldots, A^{l_i} f(i) : i \in \Omega\}$ in a stable way.

Note that, in Problem 2.1.1, we allow l_i to be finite or infinite. Note also that, Problem 2.1.1 is not the most general problem since the way it is stated implies that $\Omega = \Omega_0$ and $\Omega_n = \{i \in \Omega_0 : l_i \geq n\}$. Thus, an underlying assumption is that $\Omega_{n+1} \subseteq \Omega_n$ for all $n \geq 0$. For each $i \in \Omega$, let S_i be the operator from $\mathscr{H} = \ell^2(I)$ to $\mathscr{H}_i = \ell^2(\{0, \ldots, l_i\})$, defined by $S_i f = (A^j f(i))_{j=0,\ldots,l_i}$ and define S to be the operator $S = S_0 \oplus S_1 \oplus \ldots$

Then f can be recovered from $Y = \{f(i), Af(i), \ldots, A^{l_i} f(i) : i \in \Omega\}$ in a stable way if and only if there exist constants $c_1, c_2 > 0$ such that

$$c_1 \|f\|_2^2 \leq \|\mathscr{S} f\|_2^2 = \sum_{i \in \Omega} \|S_i f\|_2^2 \leq c_2 \|f\|_2^2. \tag{2.1}$$

Using the standard basis $\{e_i\}$ for $\ell^2(I)$, we obtain from (2.1) that

$$c_1 \|f\|_2^2 \leq \sum_{i \in \Omega} \sum_{j=0}^{l_i} |\langle f, A^{*j} e_i \rangle|^2 \leq c_2 \|f\|_2^2.$$

Thus we get

Lemma 2.1.2. *Every $f \in \ell^2(I)$ can be recovered from the measurements set*

$$Y = \{f(i), Af(i), \ldots, A^{l_i} f(i) : i \in \Omega\}$$

in a stable way if and only if the set of vectors $\{A^{*j} e_i : \ i \in \Omega,\ j = 0, \ldots, l_i\}$ is a frame for $\ell^2(I)$.

2.2 Contribution and Organization

In section 2.3 we present the results for the finite dimensional case. Specifically, Subsection 2.3.1 concerns the special case of diagonalizable operators acting on vectors in $\mathbb{C}^d$. This case is treated first in order to give some intuition about the general theory. For example, Theorem 2.3.2 explains the reconstruction properties for the examples below: Consider the following two matrices acting on $\mathbb{C}^5$.

$$P = \begin{pmatrix} 9/2 & 1/2 & -7 & 5 & -3 \\ 15/2 & 3/2 & -11 & 5 & -7 \\ 5 & 0 & -7 & 5 & -5 \\ 4 & 0 & -4 & 3 & -4 \\ 1/2 & 1/2 & -1 & 0 & 1 \end{pmatrix} \quad Q = \begin{pmatrix} 3/2 & -1/2 & 2 & 0 & 1 \\ 1/2 & 5/2 & 0 & 0 & -1 \\ 0 & 0 & 3 & 0 & 0 \\ 1 & 0 & -1 & 3 & -1 \\ -1/2 & -1/2 & 1 & 0 & 3 \end{pmatrix}.$$

For the matrix P, Theorem 2.3.2 shows that any $f \in \mathbb{C}^5$ can be recovered from the data sampled at the single "spacial" point $i = 2$, i.e., from

$$Y = \{f(2), Pf(2), P^2 f(2), P^3 f(2), P^4 f(2)\}.$$

However, if $i = 3$, i.e., $Y = \{f(3), Pf(3), P^2 f(3), P^3 f(3), P^4 f(3)\}$ the information is not sufficient to determine f. In fact if we do not sample at $i = 1$, or $i = 2$, the only way to recover any $f \in \mathbb{C}^5$ is to sample at all the remaining "spacial" points $i = 3, 4, 5$. For example,

$Y = \{f(i), Pf(i) : i = 3,4,5\}$ is enough data to recover f, but $Y = \{f(i), Pf(i), ..., P^L f(i) : i = 3,4\}$, is not enough information no matter how large L is.

For the matrix Q, Theorem 2.3.2 implies that it is not possible to reconstruct $f \in \mathbb{C}^5$ if the number of sampling points is less than 3. However, we can reconstruct any $f \in \mathbb{C}^5$ from the data

$$\begin{aligned} Y = &\{f(1), Qf(1), Q^2 f(1), Q^3 f(1), Q^4 f(1), \\ &f(2), Qf(2), Q^2 f(2), Q^3 f(2), Q^4 f(2), \\ &f(4), Qf(4)\}. \end{aligned}$$

Yet, it is not possible to recover f from the set $Y = \{Q^l f(i) : i = 1,2,3, l = 0, \ldots, L\}$ for any L. Theorem 2.3.2 gives all the sets Ω such that any $f \in \mathbb{C}^5$ can be recovered from $Y = \{A^l f(i) : i \in \Omega, l = 0, ...l_i\}$.

In subsection 2.3.2 Problem 2.1.1 is solved for the general case in $\mathbb{C}^d$, and Corollary 2.3.7 elucidates the example below: Consider

$$R = \begin{pmatrix} 0 & -1 & 4 & -1 & 2 \\ 2 & 1 & -2 & 1 & -2 \\ -1/2 & -1/2 & 3 & 0 & 1 \\ 1/2 & -1/2 & 0 & 2 & 0 \\ -1/2 & -1/2 & 2 & -1 & 2 \end{pmatrix}.$$

Then, Corollary 2.3.7 shows that Ω must contain at least two "spacial" sampling points for the recovery of functions from their time-space samples to be feasible. For example, if $\Omega = \{1,3\}$, then $Y = \{R^l f(i) : i \in \Omega, l = 0, \ldots, L\}$ is enough recover $f \in \mathbb{C}^5$. However, if Ω is changed to $\Omega = \{1,2\}$, then $Y = \{R^l f(i) : i \in \Omega, l = 0, \ldots, L\}$ does not provide enough information.

The dynamical sampling problem in infinite dimensional separable Hilbert spaces is

studied in Section 2.4. For this case, we restrict ourselves to certain classes of self adjoint operators in $\ell^2(\mathbb{N})$. In light of Lemma 2.1.2, in Subsection 2.4.1, we characterize the sets $\Omega \subseteq \mathbb{N}$ such that $\mathscr{F}_\Omega = \{A^j e_i : i \in \Omega, j = 0, \ldots, l_i\}$ is complete in $\ell^2(\mathbb{N})$ (Theorem 2.4.3). However, we also show that if Ω is a finite set, then $\{A^j e_i : i \in \Omega, j = 0, \ldots, l_i\}$ is never a basis (see Theorem 2.4.8). It turns out that the obstruction to being a basis is redundancy. This fact is proved using the beautiful Müntz-Szász Theorem 2.4.5 below.

Although $\mathscr{F}_\Omega = \{A^j e_i : i \in \Omega, j = 0, \ldots, l_i\}$ cannot be a basis, it should be possible that $\mathscr{F}_\Omega$ is a frame for sets $\Omega \subseteq \mathbb{N}$ with finite cardinality. It turns out however, that except for special cases, if Ω is a finite set, then $\mathscr{F}_\Omega$ is not a frame for $\ell^2(\mathbb{N})$.

If Ω consists of a single vector, we are able to characterize completely when $\mathscr{F}_\Omega$ is a frame for $\ell^2(\mathbb{N})$ (Theorem 2.4.16), by relating our problem to a theorem by Carleson on interpolating sequences in the Hardy spaces $H^2(\mathbb{D})$.

2.3 Finite Dimensional Case

In this section we will address the finite dimensional case. That is, our evolution operator is a matrix A acting on the space $\mathbb{C}^d$ and $I = \{1, \ldots, d\}$. Thus, given A, our goal is to find necessary and sufficient conditions on the set of indices $\Omega \subseteq I$ and the numbers $\{l_i\}_{i \in \Omega}$ such that every vector $f \in \mathbb{C}^d$ can be recovered from the samples $\{A^j f(i) : i \in \Omega,\ j = 0, \ldots, l_i\}$ or equivalently (using Lemma 2.1.2), the set of vectors

$$\{A^{*j} e_i : i \in \Omega,\ j = 0, \ldots, l_i\} \text{ is a frame of } \mathbb{C}^d. \tag{2.2}$$

(Note that this implies that we need at least d space-time samples to be able to recover the vector f).

The problem can be further reduced as follows: Let B be any invertible matrix with complex coefficients, and let Q be the matrix $Q = BA^*B^{-1}$, so that $A^* = B^{-1}QB$. Let b_i denote the ith column of B. Since a frame is transformed to a frame by invertible linear

operators, condition (2.2) is equivalent to $\{Q^j b_i : i \in \Omega,\ j = 0, \ldots, l_i\}$ being a frame of $\mathbb{C}^d$.

This allows us to replace the general matrix A^* by a possibly simpler matrix and we have:

Lemma 2.3.1. *Every $f \in \mathbb{C}^d$ can be recovered from the measurement set $Y = \{A^j f(i) : i \in \Omega, j = 0, \ldots, l_i\}$ if and only if the set of vectors $\{Q^j b_i : i \in \Omega,\ j = 0, \ldots, l_i\}$ is a frame for $\mathbb{C}^d$.*

We begin with the simpler case when A^* is a diagonalizable matrix.

2.3.1 Diagonalizable Transformations

Let $A \in \mathbb{C}^{d \times d}$ be a matrix that can be written as $A^* = B^{-1}DB$ where D is a diagonal matrix of the form

$$D = \begin{pmatrix} \lambda_1 I_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n I_n \end{pmatrix}. \tag{2.3}$$

In (2.3), I_k is an $h_k \times h_k$ identity matrix, and $B \in \mathbb{C}^{d \times d}$ is an invertible matrix. Thus A^* is a diagonalizable matrix with distinct eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$.

Using Lemma 2.3.1 and $Q = D$, Problem 2.1.1 becomes the problem of finding necessary and sufficient conditions on vectors b_i and numbers l_i, and the set $\Omega \subseteq \{1, \ldots, m\}$ such that the set of vectors $\{D^j b_i : i \in \Omega,\ j = 0, \ldots, l_i\}$ is a frame for $\mathbb{C}^d$. Recall that the *Q-annihilator q_b^Q of a vector b* is the monic polynomial of smallest degree, such that $q_b^Q(Q)b \equiv 0$. Let P_j denote the orthogonal projection in $\mathbb{C}^d$ onto the eigenspace of D associated to the eigenvalue λ_j. Then we have:

Theorem 2.3.2. *Let $\Omega \subseteq \{1, \ldots, d\}$ and $\{b_i : i \in \Omega\}$ vectors in $\mathbb{C}^d$. Let D be a diagonal matrix and r_i the degree of the D-annihilator of b_i. Set $l_i = r_i - 1$. Then $\{D^j b_i : i \in \Omega, j =$*

$0,\dots,l_i\}$ *is a frame of* $\mathbb{C}^d$ *if and only if* $\{P_j(b_i) : i \in \Omega\}$ *form a frame of* $P_j(\mathbb{C}^d)$, $j = 1,\dots,n$.

As a corollary, using Lemma 2.3.1 we get

Theorem 2.3.3. *Let* $A^* = B^{-1}DB$, *and let* $\{b_i : i \in \Omega\}$ *be the column vectors of* B *whose indices belong to* Ω. *Let* r_i *be the degree of the D-annihilator of* b_i *and let* $l_i = r_i - 1$. *Then* $\{A^{*j}e_i : \ i \in \Omega,\ j = 0,\dots,l_i\}$ *is a frame of* $\mathbb{C}^d$ *if and only if* $\{P_j(b_i) : i \in \Omega\}$ *form a frame of* $P_j(\mathbb{C}^d)$, $j = 1,\dots,n$.

Equivalently, any vector $f \in \mathbb{C}^d$ *can be recovered from the samples*

$$Y = \{f(i), Af(i), \dots, A^{l_i}f(i) : i \in \Omega\}$$

if and only if $\{P_j(b_i) : i \in \Omega\}$ *form a frame of* $P_j(\mathbb{C}^d)$, $j = 1,\dots,n$.

Example 2.2 in [6] can be derived from Theorem 2.3.3 when all the eigenvalues have multiplicity 1, and when there is a single sampling point at location i.

Note that, in the previous Theorem, the number of time-samples l_i depends on the sampling point i. If instead the number of time-samples L is the same for all $i \in \Omega$, (note that $L \geq max\{l_i : i \in \Omega\}$ is an obvious choice, but depending on the vectors b_i it may be possible to choose $L \leq min\{l_i : i \in \Omega\}$), then we have the following Theorems (see Figure **??**)

Theorem 2.3.4. *Let* D *be a diagonal matrix,* $\Omega \subseteq \{1,\dots,d\}$ *and* $\{b_i : i \in \Omega\}$ *be a set of vectors in* $\mathbb{C}^d$ *such that* $\{P_j(b_i) : i \in \Omega\}$ *form a frame of* $P_j(\mathbb{C}^d)$, $j = 1,\dots,n$. *Let* L *be any fixed integer, then* $E = \bigcup_{\{i\in\Omega : b_i \neq 0\}} \{b_i, Db_i, \dots, D^L b_i\}$ *is a frame of* $\mathbb{C}^d$ *if and only if* $\{D^{L+1}b_i, : i \in \Omega\} \subseteq span(E)$.

Proof. Note that if $\{D^{L+1}b_i : i \in \Omega\} \subseteq span(E)$ then $D(span(E)) \subseteq span(E)$. Therefore by Theorem 2.3.3, E is a frame of $\mathbb{C}^d$. □

As a corollary, for our original problem 2.1.1 we obtain

Theorem 2.3.5. *Let $A^* = B^{-1}DB$, L be any fixed integer, and let $\{b_i : i \in \Omega\}$ be a set of vectors in $\mathbb{C}^d$ such that $\{P_j(b_i) : i \in \Omega\}$ form a frame of $P_j(\mathbb{C}^d)$, $j = 1,\dots,n$. Then $\{A^{*j}e_i : i \in \Omega,\ j = 0,\dots,L\}$ is a frame of $\mathbb{C}^d$ if and only if $\{D^{L+1}b_i : i \in \Omega\} \subseteq span\big(\{D^j b_i : i \in \Omega\ , j = 0,\dots,L\}\big)$.*

Equivalently any $f \in \mathbb{C}^d$ can be recovered from the samples

$$Y = \{f(i), Af(i), A^2 f(i), \dots, A^L f(i) : i \in \Omega\},$$

if and only if $\{D^{L+1}b_i : i \in \Omega\} \subseteq span\big(\{D^j b_i : i \in \Omega\ , j = 0,\dots,L\}\big)$.

Proof. For the proof we just apply Lemma 2.3.1 and Theorem 2.3.4. □

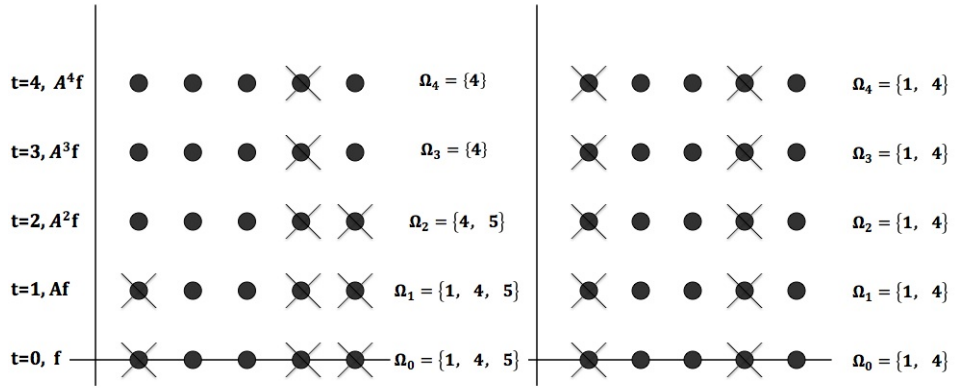

Figure 2.1: Illustration of a time-space sampling pattern. Crosses correspond to time-space sampling points. Left panel: $\Omega = \Omega_0 = \{1,4,5\}$. $l_1 = 1, l_4 = 4, l_5 = 3$. Right panel: $\Omega = \Omega_0 = \{1,4\}$. $L = 4$.

A special case of Theorem 2.3.5 is [6, Theorem 3.2]. There, since the operator A is a convolution operator in $\ell^2(\mathbb{Z}_d) \approx \mathbb{C}^d$, the matrix B is the Fourier matrix whose columns consist of the discrete, complex exponentials. The set Ω consists of the union of a uniform grid $m\mathbb{Z}_d$ and an extra sampling set Ω_0. In [6, Theorem 3.2] L can be chosen to be any number larger than m.

Theorems 2.3.3 and 2.3.5 will be consequences of our general results but we state them here to help the comprehension of the general results below.

2.3.2 General Linear Transformations

For a general matrix we will need to use the reduction to its Jordan form. To state our results in this case, we need to introduce some notations and describe the general Jordan form of a matrix with complex entries. (For these and other results about matrix or linear transformation decompositions see for example [29].)

A matrix J is in Jordan form if

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_n \end{pmatrix}. \tag{2.4}$$

In (2.4), for $s = 1, \ldots, n$, $J_s = \lambda_s I_s + N_s$ where I_s is an $h_s \times h_s$ identity matrix, and N_s is a $h_s \times h_s$ nilpotent block-matrix of the form:

$$N_s = \begin{pmatrix} N_{s1} & 0 & \cdots & 0 \\ 0 & N_{s2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_{s\gamma_s} \end{pmatrix} \tag{2.5}$$

where each N_{si} is a $t_i^s \times t_i^s$ cyclic nilpotent matrix,

$$N_{si} \in \mathbb{C}^{t_i^s \times t_i^s}, \quad N_{si} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{2.6}$$

with $t_1^s \geq t_2^s \geq \ldots$, and $t_1^s + t_2^s + \cdots + t_{\gamma_s}^s = h_s$. Also $h_1 + \cdots + h_n = d$. The matrix J has d

rows and distinct eigenvalues $\lambda_j, j = 1, \ldots, n$.

Let k_j^s denote the index corresponding to the first row of the block N_{sj} from the matrix J, and let $e_{k_j^s}$ be the corresponding element of the standard basis of $\mathbb{C}^d$. (That is a cyclic vector associated to that block). We also define $W_s := \text{span}\{e_{k_j^s} : j = 1, \ldots, \gamma_s\}$, for $s = 1, \ldots, n$, and P_s will again denote the orthogonal projection onto W_s. Finally, recall that the *J annihilator* q_b^J *of a vector b* is the monic polynomial of smallest degree, such that $q_b^J(J)b \equiv 0$. Using the notations and definitions above we can state the following theorem:

Theorem 2.3.6. *Let J be a matrix in Jordan form, as in* (2.4). *Let* $\Omega \subseteq \{1, \ldots, d\}$ *and* $\{b_i : i \in \Omega\}$ *be a subset of vectors of* $\mathbb{C}^d$, r_i *be the degree of the J-annihilator of the vector* b_i *and let* $l_i = r_i - 1$.

Then the following propositions are equivalent.

i) The set of vectors $\{J^j b_i : \ i \in \Omega, j = 0, \ldots, l_i\}$ *is a frame for* $\mathbb{C}^d$.

ii) For every $s = 1, \ldots, n$, $\{P_s(b_i), i \in \Omega\}$ *form a frame of* W_s.

Now, for a general matrix A, using Lemma 2.3.1 we can state:

Corollary 2.3.7. *Let A be a matrix, such that* $A^* = B^{-1}JB$, *where* $J \in \mathbb{C}^{d\times d}$ *is the Jordan matrix for* A^*. *Let* $\{b_i : i \in \Omega\}$ *be a subset of the column vectors of B,* r_i *be the degree of the J-annihilator of the vector* b_i, *and let* $l_i = r_i - 1$.
Then, every $f \in \mathbb{C}^d$ *can be recovered from the measurement set* $Y = \{(A^j f)(i) : \ i \in \Omega, j = 0, \ldots, l_i\}$ *of* $\mathbb{C}^d$ *if and only if* $\{P_s(b_i), i \in \Omega\}$ *form a frame of* W_s.

In other words, we will be able to recover f from the measurements Y, if and only if the *Jordan*-vectors of A^* (i.e. the columns of the matrix B that reduces A^* to its Jordan form) corresponding to Ω satisfy that their projections on the spaces W_s form a frame.

Remark 2.3.8. We want to emphasize at this point, that given a matrix in Jordan form there is an obvious choice of vectors in order that their iterations give a frame of the space, (namely, the cyclic vectors $e_{k_j^s}$ corresponding to each block). However, we are dealing here

with a much more difficult problem. The vectors b_i are given beforehand, and we need to find conditions in order to decide if their iterations form a frame.

The following theorem is just a statement about replacing the *optimal* iteration of each vector b_i by any fixed number of iterations. The idea is, that we iterate a fixed number of times L but we do not need to know the degree r_i of the J-annihilator for each b_i. Clearly, if $L \geq \max\{r_i - 1 : i \in \Omega\}$ then we can always recover any f from Y. But the number of time iterations L may be smaller than any $r_i - 1$, $i \in \Omega$. In fact, for practical purposes it might be better to iterate, than to try to figure out which is the degree of the annihilator for b_i.

Theorem 2.3.9. *Let $J \in \mathbb{C}^{d\times d}$ be a matrix in Jordan form (see* (2.4)*). Let $\Omega \subseteq \{1,\dots,d\}$, and let $\{b_i : i \in \Omega\}$ be a set of vectors in $\mathbb{C}^d$, such that for each $s = 1,\dots,n$ the projections $\{P_s(b_i) : i \in \Omega\}$ onto W_s form a frame of W_s. Let L be any fixed integer, then $E = \bigcup_{\{i\in\Omega : b_i \neq 0\}} \{b_i, Jb_i, \dots, J^L b_i\}$ is a frame of $\mathbb{C}^d$ if and only if $\{J^{L+1}b_i : i \in \Omega\} \subseteq span(E)$.*

As a corollary we immediately get the solution to Problem 2.1.1 in finite dimensions.

Corollary 2.3.10. *Let $\Omega \subseteq I$, $A^* = B^{-1}JB$, and L be any fixed integer. Assume that $\{P_s(b_i) : i \in \Omega\}$ form a frame of W_s and set $E = \{J^s b_i : i \in \Omega, s = 0,\dots,L,\}$. Then any $f \in \mathbb{C}^d$ can be recovered from the samples $Y = \{f(i), Af(i), A^2 f(i), \dots, A^L f(i) : i \in \Omega\}$, if and only if $\{J^{L+1}b_i : i \in \Omega\} \subseteq span(E\})$.*

2.3.3 Proofs

In order to introduce some needed notations, we first recall the standard decomposition of a linear transformation acting on a finite dimensional vector space that produces a basis for the Jordan form.

Let V be a finite dimensional vector space of dimension d over $\mathbb{C}$ and let $T : V \longrightarrow V$ be a linear transformation. The characteristic polynomial of T factorizes as $\chi_T(x) = (x - \lambda_1)^{h_1} \dots (x - \lambda_n)^{h_n}$ where $h_i \geq 1$ and $\lambda_1, \dots, \lambda_n$ are distinct elements of $\mathbb{C}$. The minimal

polynomial of T will be then $m_T(x) = (x-\lambda_1)^{r_1}\dots(x-\lambda_n)^{r_n}$ with $1 \leq r_i \leq h_i$ for $i = 1,\dots,n$. By the primary decomposition theorem, the subspaces $V_s = \mathrm{Ker}(T-\lambda_s I)^{r_s}$, $s = 1,\dots,n$ are invariant under T (i.e. $T(V_s) \subseteq V_s$) and we have also that $V = V_1 \oplus \cdots \oplus V_n$.

Let T_s be the restriction of T to V_s. Then, the minimal polynomial of T_s is $(x-\lambda_s)^{r_s}$, and $T_s = N_s + \lambda_s I_s$, where N_s is nilpotent of order r_s and I_s is the identity operator on V_s. Now for each s we apply the cyclic decomposition to N_s and the space V_s to obtain:

$$V_s = V_{s1} \oplus \cdots \oplus V_{s\gamma_s}$$

where each V_{sj} is invariant under N_s, and the restriction operator N_{sj} of N_s to V_{sj} is a *cyclic nilpotent* operator on V_{sj}.

Finally, let us fix for each j a cyclic vector $w_{sj} \in V_{sj}$ and define the subspace $W_s = \mathrm{span}\{w_{s1}\dots w_{s\gamma_s}\}$, $W = W_1 \oplus \cdots \oplus W_n$ and let P_{W_s} be the projection onto W_s, with $I_W = P_{W_1} + \cdots + P_{W_n}$.

With this notation we can state the main theorem of this section:

Theorem 2.3.11. *Let $\{b_i : i \in \Omega\}$ be a set of vectors in V. If the set $\{P_{W_s} b_i : i \in \Omega\}$ is complete in W_s for each $s = 1,\dots,n$, then the set $\{b_i, Tb_i, \dots, T^{l_i} b_i : i \in \Omega\}$ is a frame of V, where r_i is the degree of the T-annihilator of b_i and $l_i = r_i - 1$.*

To prove Theorem 2.3.11, we will first concentrate on the case where the transformation T has minimal polynomial consisting of a unique factor, i.e. $m_T(x) = (x-\lambda)^r$, so that $T = \lambda I_d + N$, and $N^r = 0$ but $N^{r-1} \neq 0$.

2.3.4 Case $T = \lambda I_d + N$

Remark 2.3.12. It is not difficult to see that, in this case, given some $L \in \mathbb{N}$, $\{T^j b_i : i \in \Omega, j = 0,\dots,L\}$ is a frame for V if and only if $\{N^j b_i : i \in \Omega, j = 0,\dots,L\}$ is a frame for V. In addition, since $N^r b_i = 0$ we need only to iterate to $r-1$. In fact, we only need to iterate each b_i to $l_i = r_i - 1$ where r_i is the degree of the N annihilator of b_i.

Definition 2.3.13. A matrix $A \in \mathbb{C}^{d\times d}$ is *perfect* if $a_{ii} \neq 0, i = 1,\ldots,d$ and $\det(A_i) \neq 0, i = 1,\ldots,d$ where $A_s \in \mathbb{C}^{s\times s}$ is the submatrix of A, $A_s = \{a_{i,j}\}_{i,j=1,\ldots,s}$.

We need the following lemma that is straightforward to prove.

Lemma 2.3.14. *Let $A \in \mathbb{C}^{d\times d}$ be an invertible matrix. Then there exists a perfect matrix $B \in \mathbb{C}^{d\times d}$ that consists of row (or column) permutations of A.*

Proof. The proof is by induction on d, which is the number of rows (or columns) of the matrix. The case of $d = 1$ is obvious, so let A be an invertible $d \times d$ matrix with entries $a_{i,j}$ and assume that the lemma is true for dimension $d-1$. Let us expand the determinant of A using the last column, i.e.:

$$\det(A) = \sum_{i=1}^{d} (-1)^{i+d}\, a_{i,d}\, \det(A^{(i,d)}),$$

where $A^{(i,j)}$ denotes the $(d-1)\times(d-1)$ submatrix of A that is obtained by removing the row i and the column j from A.

Since $\det(A)$ is different from zero, there exists $i \in \{1,\ldots,d\}$ such that $a_{i,d}$ and $det(A^{(i,d)})$ are both different from zero. Let B be the matrix obtained from A by interchanging row i with row d. So the $(d-1)\times(d-1)$ submatrix B_{d-1} of B obtained by removing row d and column d from B, is invertible and the element of B, $b_{d,d} = a_{i,d}$ is not zero.

We now apply the inductive hypothesis to the matrix B_{d-1}. So there exits some permutation of the rows of B_{d-1} such that the matrix is perfect. If we apply the same permutation to the firs $d-1$ rows of B, we obtain a matrix $\tilde{B}$ such that $\tilde{B}_{d-1}$ is perfect and its (d,d)th entry is non zero. Therefore $\tilde{B}$ is perfect and has been obtained from A by permutation of the rows.

□

If N is nilpotent of order r, then there exist $\gamma \in \mathbb{N}$ and invariant subspaces $V_i \subseteq V$,

$i = 1, \ldots, \gamma$ such that

$$V = V_1 \oplus \cdots \oplus V_\gamma, \quad \dim(V_j) = t_j,\ t_j \geq t_{j+1},\ j = 1, \ldots, \gamma - 1,$$

and $N = N_1 + \cdots + N_\gamma$, where $N_j = P_j N P_j$ is a cyclic nilpotent operator in V_j, $j = 1, \ldots, \gamma$. Here P_j is the projection onto V_j. Note that $t_1 + \cdots + t_\gamma = d$.

For each $j = 1, \ldots, \gamma$, let $w_j \in V_j$ be a cyclic vector for N_j. Note that the set $\{w_1, \ldots, w_\gamma\}$ is a linearly independent set.

Let $W = \text{span}\{w_1, \ldots, w_\gamma\}$. Then, we can write $V = W \oplus NW \oplus \cdots \oplus N^{r-1}W$. Furthermore, the projections P_{N^jW} satisfy $P_{N^jW}^2 = P_{N^jW}$, and $I = \sum_{j=0}^{r-1} P_{N^jW}$.

Finally, note that

$$N^s P_W = P_{N^sW} N^s. \tag{2.7}$$

With the notation above, we have the following theorem:

Theorem 2.3.15. *Let N be a nilpotent operator on V. Let $B \subseteq V$ be a finite set of vectors such that $\{P_W(b) : b \in B\}$ is complete in W. Then*

$$\bigcup_{b \in B} \left\{ b, Nb, \ldots, N^{l_b} b \right\} \quad \textit{is a frame for } V,$$

where $l_b = r_b - 1$ and r_b is the degree of the N-annihilator of b.

Proof. In order to prove Theorem 2.3.15, we will show that there exist vectors $\{b_1, \ldots, b_\gamma\}$ in B, where $\gamma = \dim(W)$, such that

$$\bigcup_{i=1}^{\gamma} \left\{ b_i, Nb_i, \ldots, N^{t_i - 1} b_i \right\} \quad \text{is a basis of } V.$$

Recall that t_i are the dimensions of V_i defined above. Since $\{P_W(b) : b \in B\}$ is complete in W and $\dim(W) = \gamma$ it is clear that we can choose $\{b_1, \ldots, b_\gamma\} \subseteq B$ such that $\{P_W(b_i) : i = 1, \ldots, \gamma\}$ is a basis of W. Since $\{w_1, \ldots, w_\gamma\}$ is also a basis of W, there exist unique scalars

$\{\theta_{i,j} : i,j = 1,\ldots,\gamma\}$ such that,

$$P_W(b_i) = \sum_{j=1}^{\gamma} \theta_{ij} w_j. \tag{2.8}$$

with the matrix $\Theta = \{\theta_{i,j}\}_{i,j=1,\ldots,\gamma}$ invertible. Thus, using Lemma 2.3.14 we can relabel the indices of $\{b_i\}$ in such a way that Θ is *perfect*. Therefore, without loss of generality, we can assume that $\{b_1,\ldots,b_\gamma\}$ are already in the right order, so that Θ is perfect.

We will now prove that the d vectors $\left\{b_i, Nb_i, \ldots, N^{t_i-1}b_i\right\}_{i=1,\ldots,\gamma}$ are linearly independent. For this, assume that there exist scalars α_j^s such that

$$0 = \sum_{j=1}^{\gamma} \alpha_j^0 b_j + \sum_{j=1}^{p_1} \alpha_j^1 N b_j + \cdots + \sum_{j=1}^{p_{r-1}} \alpha_j^{r-1} N^{r-1} b_j, \tag{2.9}$$

where $p_s = \max\{j : t_j > s\} = \dim N^s W, s = 1,\ldots,r-1$ (note that $p_s \geq 1$, since $N^{r-1}b_1 \neq 0$).

Note that since $V = W \oplus NW \oplus \cdots \oplus N^{r-1}W$, for any vector $x \in V$, $P_W(Nx) = 0$. Therefore, if we apply P_W on both sides of (2.9), we obtain

$$\sum_{j=1}^{\gamma} \alpha_j^0 P_W b_j = 0.$$

Since $\{P_W b_i : i = 1,\ldots,\gamma\}$ are linearly independent, we have $\alpha_j^0 = 0,\ j = 1,\ldots,\gamma$. Hence, if we now apply P_{NW} to (2.9), we have as before that

$$\sum_{j=1}^{p_1} \alpha_j^1 P_{NW} N b_j = 0.$$

Using the conmutation property of the projection, (2.7), we have

$$\sum_{j=1}^{p_1} \alpha_j^1 N P_W b_j = 0.$$

In matrix notation, this is

$$[\alpha_1^1 \dots \alpha_{p_1}^1]\Theta_{p_1} \begin{bmatrix} Nw_1 \\ \vdots \\ Nw_{p_1} \end{bmatrix} = 0.$$

Note that by definition of p_1, $Nw_1, \dots, Nw_{p_1}$ span NW, and since the dimension of NW is exactly p_1, $Nw_1, \dots, Nw_{p_1}$ are linearly independent vectors. Therefore $[\alpha_1^1 \dots \alpha_{p_1}^1]\Theta_{p_1} = 0$. Since Θ is perfect, $[\alpha_1^1 \dots \alpha_{p_1}^1] = [0 \dots 0]$. Iterating the above argument, the Theorem follows. □

Proof of Theorem 2.3.11 .

We will prove the case when the minimal polynomial has only two factors. The general case follows by induction.

That is, let $T : V \to V$ be a linear transformation with characteristic polynomial of the form $\chi_T(x) = (x-\lambda_1)^{h_1}(x-\lambda_2)^{h_2}$. Thus, $V = V_1 \oplus V_2$ where V_1, V_2 are the subspaces associated to each factor, and $T = T_1 \oplus T_2$. In addition, $W = W_1 \oplus W_2$ where W_1, W_2 are the subspaces of the cyclic vectors from the cyclic decomposition of N_1 with respect of V_1 and of N_2 with respect to V_2.

Let $\{b_i : i \in \Omega\}$ be vectors in V that satisfy the hypothesis of the Theorem. For each b_i we write $b_i = c_i + d_i$ with $c_i \in V_1$ and $d_i \in V_2$, $i \in \Omega$. Let r_i, m_i and n_i be the degrees of the annihilators $q_{b_i}^T$, $q_{c_i}^{T_1}$ and $q_{d_i}^{T_2}$, respectively. By hypothesis $\{P_{W_1} c_i : i \in \Omega\}$ and $\{P_{W_2} d_i : i \in \Omega\}$ are complete in W_1 and W_2, respectively. Hence, applying Theorem 2.3.15 to N_1 and N_2 we conclude that $\bigcup_{i\in\Omega}\{T_1^j c_i, j = 0, 1, \dots m_i - 1\}$ is complete in V_1, and that $\bigcup_{i\in\Omega}\{T_2^j d_i, j = 0, 1, \dots n_i - 1\}$ is complete in V_2.

We will now need a Lemma: (Recall that q_b^T is the T-annihilator of the vector b)

Lemma 2.3.16. *Let T be as above, and $V = V_1 \oplus V_2$. Given $b \in V$, $b = c + d$ then $q_b^T = q_c^{T_1} q_d^{T_2}$ where $q_c^{T_1}$ and $q_d^{T_2}$ are coprime. Further let $u \in V_2$, $u = q_c^{T_1}(T_2)d$. Then $q_u^{T_2}$ coincides*

with $q_d^{T_2}$.

Proof. The fact that $q_b^T = q_c^{T_1} q_d^{T_2}$ with coprime $q_c^{T_1}$ and $q_d^{T_2}$ is a consequence of the decomposition of T.

Now, by definition of $q_u^{T_2}$ we have that

$$0 = q_u^{T_2}(T_2)(u) = q_u^{T_2}(T_2)(q_c^{T_1}(T_2)d) = (q_u^{T_2} q_c^{T_1})(T_2)d.$$

Thus, $q_d^{T_2}$ has to divide $q_u^{T_2} \cdot q_c^{T_1}$, but since $q_d^{T_2}$ is coprime with $q_c^{T_1}$, we conclude that

$$q_d^{T_2} \quad \text{divides} \quad q_u^{T_2}. \tag{2.10}$$

On the other hand

$$\begin{aligned} 0 &= q_d^{T_2}(T_2)(d) = q_c^{T_1}(T_2)(q_d^{T_2}(T_2)d) = (q_c^{T_1} q_d^{T_2})(T_2)d \\ &= (q_d^{T_2} q_c^{T_1})(T_2)d = q_d^{T_2}(T_2)(q_c^{T_1}(T_2)d) = q_d^{T_2}(T_2)(u), \end{aligned}$$

and therefore

$$q_u^{T_2} \quad \text{divides} \quad q_d^{T_2}. \tag{2.11}$$

From (2.10) and (2.11) we obtain $q_d^{T_2} = q_u^{T_2}$. □

Now, we continue with the proof of the Theorem. Recall r_i, m_i and n_i be the degrees of $q_{b_i}^T$, $q_{c_i}^{T_1}$ and $q_{d_i}^{T_2}$, respectively, and let $l_i = r_i - 1$. Also note that by Lemma 2.3.16 $r_i = m_i + n_i$. In order to prove that the set $\{b_i, Tb_i, \ldots, T^{l_i} b_i : i \in \Omega\}$ is complete in V, we will replace this set with a new one in such a way that the dimension of the span does not change.

For each $i \in \Omega$, let $u_i = q_{c_i}^{T_1}(T_2)d_i$. Now, for a fixed i we leave the vectors $b_i, Tb_i, \ldots, T^{m_i-1} b_i$ unchanged, but for $s = 0, \ldots, n_i - 1$ we replace the vectors $T^{m_i+s} b_i$ by the vectors $T^{m_i+s} b_i + \beta_s(T) b_i$ where β_s is the polynomial $\beta_s(x) = x^s q_{c_i}^{T_1}(x) - x^{m_i+s}$.

Note that $\text{span}\{b_i, Tb_i, \ldots, T^{m_i+s}b_i\}$ remains unchanged, since $\beta_s(T)b_i$ is a linear combination of the vectors $\{T^s b_i, \ldots, T^{m_i+s-1}b_i\}$.

Now we observe that:

$$T^{m_i+s}b_i + \beta_s(T)b_i = \left[T_1^{m_i+s}c_i + \beta_s(T_1)c_i\right] + \left[T_2^{m_i+s}d_i + \beta_s(T_2)d_i\right].$$

The first term of the sum on the right hand side of the equation above is in V_1 and the second in V_2. By definition of β_s we have:

$$T_1^{m_i+s}c_i + \beta_s(T_1)c_i = T_1^{m_i+s}c_i + T_1^s q_{c_i}^{T_1}(T_1)c_i - T_1^{m_i+s}c_i = T_1^s q_{c_i}^{T_1}(T_1)c_i = 0,$$

and

$$T_2^{m_i+s}d_i + \beta_s(T_2)d_i = T_2^s q_{c_i}^{T_1}(T_2)(d_i) = T_2^s u_i.$$

Thus, for each $i \in \Omega$, the vectors $\{b_i, \ldots, T^{l_i}b_i\}$ have been replaced by the vectors

$$\{b_i, \ldots, T^{m_i-1}b_i, u_i, \ldots, T^{n_i-1}u_i\}$$

and both sets have the same span.

To finish the proof we only need to show that the new system is complete in V.

Using Lemma 2.3.16, we have that for each $i \in \Omega$,

$$\dim(\text{span}\{u_i, \ldots, T_2^{n_i-1}u_i\}) = \dim(\text{span}\{d_i, \ldots, T_2^{n_i-1}d_i\}) = n_i,$$

and since each $T_2^s u_i \in \text{span}\{d_i, \ldots, T_2^{n_i-1}d_i\}$ we conclude that

$$\text{span}\{u_i, \ldots, T_2^{n_i-1}u_i : i \in \Omega\} = \text{span}\{d_i, \ldots, T_2^{n_i-1}d_i : i \in \Omega\}. \tag{2.12}$$

Now assume that $x \in V$ with $x = x_1 + x_2,\ x_i \in V_i$. Since by hypothesis span$\{c_i, \ldots, T_1^{m_i-1} c_i : i \in \Omega\}$ is complete in V_1, we can write

$$x_1 = \sum_{i \in \Omega} \sum_{j=0}^{m_i-1} \alpha_j^i T_1^j c_i, \tag{2.13}$$

for same scalars α_j^i, and therefore,

$$\sum_{i \in \Omega} \sum_{j=0}^{m_i-1} \alpha_j^i T^j b_i = x_1 + \sum_{i \in \Omega} \sum_{j=0}^{m_i-1} \alpha_j^i T_2^j d_i = x_1 + \tilde{x}_2, \tag{2.14}$$

since $\sum_{i \in \Omega} \sum_{j=0}^{m_i-1} \alpha_j^i T_2^j d_i = \tilde{x}_2$ is in V_2 by the invariance of V_2 by T. Since by hypothesis $\{T_2^j d_i : i \in \Omega, \quad j = 1, \ldots, n_i - 1\}$ is complete in V_2, by equation (2.12), $\{T_2^j u_i : i \in \Omega, \quad j = 1, \ldots, n_i - 1\}$ is also complete in V_2, and therefore there exist scalars β_j^i,

$$x_2 - \tilde{x}_2 = \sum_{i \in \Omega} \sum_{j=0}^{n_i-1} \beta_j^i T_2^j u_i,$$

and so

$$x = \sum_{i \in \Omega} \sum_{j=0}^{m_i-1} \alpha_j^i T^j b_i + \sum_{i \in \Omega} \sum_{j=0}^{n_i-1} \beta_j^i T_2^j u_i,$$

which completes the proof of Theorem 2.3.11 for the case of two coprime factors in the minimal polynomial of J. The general case of more factors follows by induction adapting the previous argument.

□

Theorem 2.3.6 and Theorem 2.3.9 and its corollaries are easy consequences of Theorem 2.3.11.

Proof of Theorem 2.3.9. Note that if $\{J^{L+1} b_i : i \in \Omega\} \subseteq \text{span}(E)$, then $\{J^{L+2} b_i : i \in \Omega\} \subseteq \text{span}(E)$ as well. Continuing in this way, it follows that for each $i \in \Omega$, $span(E)$ contains all the powers $J^j b_i$ for any j. Therefore, using Theorem 2.3.6, it follows that $span(E)$ contains a frame of $\mathbb{C}^d$, so that, $span(E) = \mathbb{C}^d$ and E is a frame of $\mathbb{C}^d$. The converse is obvious. □

The proof of Theorem 2.3.5 uses a similar argument.

Although Theorem 2.3.2 is a direct consequence of Theorem 2.3.6, we will give a simpler proof for this case.

Proof of Theorem 2.3.2.

Let $\{P_j(b_i) : i \in \Omega\}$ form a frame of $P_j(\mathbb{C}^d)$, for each $j = 1, \ldots, n$. Since we are working with finite dimensional spaces, to show that $\{D^j b_i : \ i \in \Omega,\ j = 0, \ldots, l_i\}$ is a frame of $\mathbb{C}^d$, all we need to show is that it is complete in $\mathbb{C}^d$. Let x be any vector in $\mathbb{C}^d$, then $x = \sum_{j=1}^{n} P_j x$. Assume that $\langle D^l b_i, x\rangle = 0$ for all $i \in \Omega$ and $l = 0, \ldots, l_i$. Since $l_i = r_i - 1$, where r_i is the degree of the D-annihilator of b_i, we have that $\langle D^l b_i, x\rangle = 0$ for all $i \in \Omega$ and $l = 0, \ldots, d$. In particular, since $n \leq d$, $\langle D^l b_i, x\rangle = 0$ for all $i \in \Omega$ and $l = 0, \ldots, n$. Then

$$\langle D^l b_i, x\rangle = \sum_{j=1}^{n} \langle D^l b_i, P_j x\rangle = \sum_{j=1}^{n} \lambda_j^l \langle P_j b_i, P_j x\rangle = 0, \tag{2.15}$$

for all $i \in \Omega$ and $l = 0, \ldots, n$. Let z_i be the vector $\big(\langle P_j b_i, P_j x\rangle\big) \in \mathbb{C}^n$. Then for each i, (2.15) can be written in matrix form as $V z_i = 0$ where V is the $n \times n$ Vandermonde matrix

$$V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix}, \tag{2.16}$$

which is invertible since, by assumption, the λ_js are distinct. Thus, $z_i = 0$. Hence, for each j, we have that $\langle P_j b_i, P_j x\rangle = 0$ for all $i \in \Omega$. Since $\{P_j(b_i) : i \in \Omega\}$ form a frame of $P_j(\mathbb{C}^d)$, $P_j x = 0$. Hence, $P_j x = 0$ for $j = 1, \ldots, n$ and therefore $x = 0$.

□

Remark 2.3.17. Given a general linear transformation $T : V \longrightarrow V$, the cyclic decomposition theorem gives the rational form for the matrix of T in some special basis. A natural question is then if we can obtain a similar result to Theorem 2.3.11 for this decomposition.

(Rational form instead of Jordan form). The answer is no. That is, if a set of vectors b_i with $i \in \Omega$ where Ω is a finite subset of $\{1,\dots,d\}$ when projected onto the subspace generated by the cyclic vectors, is complete in this subspace, this does not necessarily imply that its iterations $T^j b_i$ are complete in V. The following example illustrates this fact for a single cyclic operator.

- Let T be the linear transformation in $\mathbb{R}^3$ given as multiplication by the following matrix M.

$$M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

  The matrix M is in rational form with just one cyclic block. The vector $e_1 = (1,0,0)$ is cyclic for M. However it is easy to see that there exists a vector $b = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in $\mathbb{R}^3$ such that $P_W(b) = x_1 \neq 0$, (here W is span$\{e_1\}$), but $\{b, Mb, M^2 b\}$ are linearly dependent, and hence do not span $\mathbb{R}^3$. So our proof for the Jordan form uses the fact that the cyclic components in the Jordan decomposition are nilpotent!

2.4 Infinite Dimensional Case

In this section we consider the dynamical sampling problem in a separable Hilbert space $\mathscr{H}$, that without any loss of generality can be considered to be $\ell^2(\mathbb{N})$. The evolution operators that we will consider belong to the following class $\mathscr{A}$ of bounded self adjoint operators:

$$\mathscr{A} = \{A \in \mathscr{B}(\ell^2(\mathbb{N})) : A = A^*, \text{and there exists a basis of } \ell^2(\mathbb{N}) \text{ of eigenvectors of } A\}.$$

The notation $\mathscr{B}(\mathscr{H})$ stands for the bounded linear operators on the Hilbert space $\mathscr{H}$. So, if $A \in \mathscr{A}$ there exists an unitary operator B such that $A = B^*DB$ with $D = \sum_j \lambda_j P_j$ with pure spectrum $\sigma_p(A) = \{\lambda_j : j \in \mathbb{N}\} \subseteq \mathbb{R}$, with $\sup_j |\lambda_j| < +\infty$ and orthogonal projections $\{P_j\}$ such that $\sum_j P_j = I$ and $P_j P_k = 0$ for $j \neq k$. Note that the class $\mathscr{A}$ includes all the bounded self-adjoint compact operators.

Recall that a set $\{v_k\}$ in a Hilbert space $\mathscr{H}$ is

- **complete**, if $\langle f, v_k \rangle = 0 \; \forall k \implies f = 0$,
- **minimal** if $\forall j, v_j \notin \overline{\text{span}}\{v_k\}_{k \neq j}$,
- a **frame** if there exist constants $C_1, C_2 > 0$ such that for all $f \in \mathscr{H}$, $A\|f\|^2_{\mathscr{H}} \leq \sum_k |\langle f, v_k \rangle|^2 \leq B\|f\|^2_{\mathscr{H}}$, and
- a **Riesz basis**, if it is a basis which is also a frame.

Remark 2.4.1. Note that by the definition of $\mathscr{A}$, we have that for any $f \in \ell^2(\mathbb{N})$ and $l = 0, 1, \ldots$

$$< f, A^l e_j > = < f, B^* D^l B e_j > = < Bf, D^l b_j > \quad \text{and} \quad \|A^l\| = \|D^l\|.$$

It follows that $\mathscr{F}_\Omega = \left\{A^l e_i : \; i \in \Omega, l = 0, \ldots, l_i\right\}$ is complete, (minimal, frame) if and only if $\left\{D^l b_i : \; i \in \Omega, l = 0, \ldots, l_i\right\}$ is complete (minimal, frame).

2.4.1 Completeness

In this section, we characterize the sampling sets $\Omega \subseteq \mathbb{N}$ such that a function $f \in \ell^2(\mathbb{N})$ can be recovered from the data

$$Y = \{f(i), Af(i), A^2 f(i), \ldots, A^{l_i} f(i) : i \in \Omega\}$$

where $A \in \mathscr{A}$, and $0 \leq l_i \leq \infty$.

Definition 2.4.2. Given $A \in \mathscr{A}$, for each set Ω we consider the set of vectors $O_\Omega := \{b_j = Be_j : j \in \Omega\}$, where e_j is the jth canonical vector of $\ell^2(\mathbb{N})$. For each $b_i \in O_\Omega$ we define r_i to be the degree of the D-annihilator of b_i if such an annihilator exists, or we set $r_i = \infty$. Since B is unitary, this number r_i is also the degree of the A-annihilator of e_i. for the remainder of this paper we let $l_i = r_i - 1$. Also, for convenience of notation, let $\Omega_\infty := \{i \in \Omega : l_i = \infty\}$.

Theorem 2.4.3. *Let $A \in \mathscr{A}$ and $\Omega \subseteq \mathbb{N}$. Then the set $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ is complete in $\ell^2(\mathbb{N})$ if and only if for each j, the set $\left\{P_j(b_i) : i \in \Omega\right\}$ is complete on the range E_j of P_j.*

Remarks 2.4.4.

i) Note that Theorem 2.4.3 implies that $|\Omega| \geq \sup_j \dim(E_j)$. Thus, if some eigen-space has infinite dimension or if $\sup_j \dim(E_j) = +\infty$, then it is necessary to have infinitely many "spacial" sampling points in order to recover f. In particular if Ω is finite, a necessary condition on A in order for $\mathscr{F}_\Omega$ to be complete is that for all j, $\dim(E_j) < M < +\infty$ for some positive constant M.

ii) Theorem 2.4.3 can be extended to a larger class of operators. For example, for the class of operators $\widetilde{\mathscr{A}}$ in $\mathscr{B}(\ell^2(\mathbb{N}))$ in which $A \in \widetilde{\mathscr{A}}$ if $A = B^{-1}DB$ where with $D = \sum_j \lambda_j P_j$ with pure spectrum $\sigma_p(A) = \{\lambda_j : j \in \mathbb{N}\} \subseteq \mathbb{C}$ and orthogonal projections $\{P_j\}$ such that $\sum_j P_j = I$ and $P_j P_k = 0$ for $j \neq k$.

Proof of Theorem 2.4.3.

By Remark 2.4.1, to prove the theorem we only need to show that $\left\{D^l b_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ is complete if and only if for each j, the set $\left\{P_j(b_i) : i \in \Omega\right\}$ is complete in the range E_j of P_j.

Assume that $\left\{D^l b_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ is complete. For a fixed j, let $g \in E_j$ and assume that $< g, P_j b_i >= 0$ for all $i \in \Omega$. Then for any $l = 0, 1, \ldots, l_i$, we have

$$\lambda_j^l < g, P_j b_i >=< g, \lambda_j^l P_j b_i >=< g, P_j D^l b_i >=< g, D^l b_i >= 0.$$

Since $\{D^l b_i : i \in \Omega, l = 0, \ldots, l_i\}$ is complete in $\ell^2(\mathbb{N})$, $g = 0$. It follows that $\{P_j(b_i) : i \in \Omega\}$ is complete on the range E_j of P_j.

Now assume that $\{P_j(b_i) : i \in \Omega\}$ is complete in the range E_j of P_j. Let $S = \overline{span}\{D^l b_i : i \in \Omega, l = 0, \ldots, l_i\}$. Clearly $DS \subseteq S$. Thus S is invariant for D. Since D is self-adjoint, $S^\perp$ is also invariant for D. It follows that the orthogonal projection $P_{S^\perp}$ commutes with D. Thus $P_{S^\perp} = \sum_j P_j P_{S^\perp} P_j$. Multiplying this last expression by P_k from the right, we get that $P_{S^\perp} P_k = P_k P_{S^\perp} P_k$. Multiplying to the left, we get that $P_k P_{S^\perp} = P_k P_{S^\perp} P_k$. Hence, $P_{S^\perp}$ commutes with P_k for all k. Therefore, for each $i \in \Omega$, $0 = P_j P_{S^\perp}(b_i) = P_{S^\perp} P_j(b_i)$.

So $P_{S^\perp}$ is zero in E_j for all j (since $\{P_j(b_i) : i \in \Omega\}$ is complete in E_j). Hence $P_{S^\perp}$ is zero everywhere which implies that $S^\perp$ is the zero subspace. That is $S = \ell^2(\mathbb{N})$, and $\mathscr{F}_\Omega$ is complete which finishes the proof of the theorem.

□

2.4.2 Minimality and bases for the dynamical sampling in infinite dimensional Hilbert spaces

In this section we will show, that if $\Omega \subseteq \mathbb{N}$ is finite, and the set $\mathscr{F}_\Omega = \{A^l e_i : i \in \Omega, l = 0, \ldots, l_i\}$ is complete, then it can *never* be minimal, and hence the set $\mathscr{F}_\Omega$ is never a basis. In some sense, the set $\mathscr{F}_\Omega$ contains many "redundant vectors" which prevents it from being a basis. However, since $\mathscr{F}_\Omega$ is complete, this redundancy may help $\mathscr{F}_\Omega$ to be a frame. We will discuss this issue in the next section. For this section, we need the celebrated Müntz-Szász Theorem characterizing the sequences of monomials that are complete in $C[0,1]$ or $C[a,b]$ [26]:

Theorem 2.4.5 (Müntz-Szász Theorem)**.** *Let $0 \leq n_1 \leq n_2 \leq \ldots$ be an increasing sequence of nonnegative integers that goes to $+\infty$. Then*

1. *$\{x^{n_k}\}$ is complete in $C[0,1]$ if and only if $n_1 = 0$ and $\sum_{k=2}^{\infty} 1/n_k = \infty$.*

2. *If* $0 < a < b < \infty$, *then* $\{x^{n_k}\}$ *is complete in* $C[a,b]$ *if and only if* $\sum_{k=2}^{\infty} 1/n_k = \infty$.

We are now ready to state the main results of this section.

Theorem 2.4.6. *Let* $A \in \mathscr{A}$ *and let* Ω *be a non-empty subset of* $\mathbb{N}$. *If there exists* $b_i \in O_\Omega$ *such that* $r_i = \infty$, *then the set* $\mathscr{F}_\Omega$ *is not minimal.*

As an immediate corollary we get

Theorem 2.4.7. *Let* $A \in \mathscr{A}$ *and let* Ω *be a finite subset of* $\mathbb{N}$. *If* $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ *is complete in* $\ell^2(\mathbb{N})$, *then* $\mathscr{F}_\Omega$ *is not minimal in* $\ell^2(\mathbb{N})$.

Proof. Since $\mathscr{F}_\Omega \in \ell^2(\mathbb{N})$, there exists some b_i with $r_i = \infty$ and then Theorem 2.4.6 applies. □

Another immediate corollary is

Theorem 2.4.8. *Let* $A \in \mathscr{A}$ *and let* Ω *be a finite subset of* $\mathbb{N}$. *Then the set* $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ *is not a basis for* $\ell^2(\mathbb{N})$.

Proof. A basis is a complete set, so the result is a consequence of Theorem 2.4.7. □

Remarks 2.4.9.

1. Theorem 2.4.8 remains true for the class of operators $A \in \widetilde{\mathscr{A}}$ described in Remark 2.4.4.

2. Theorems 2.4.7 and 2.4.8 do not hold in the case of Ω being an infinite set. A trivial example is when $A = I$ is the identity matrix and $\Omega = \mathbb{N}$. A less trivial example is when $B \in \ell^2(\mathbb{Z})$ is the symmetric bi-infinite matrix with entries $B_{ii} = 1$, $B_{i(i+1)} = 1/4$ and $B_{i(i+k)} = 0$ for $k \geq 2$. Let $\Omega = 3\mathbb{Z}$ and $D_{kk} = 2$ if $k = 3\mathbb{Z}$, $D_{kk} = 1$ if $k = 3\mathbb{Z}+1$, and $D_{kk} = -1$ if $k = 3\mathbb{Z}+2$. Then $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, 2\right\}$ is a basis for $\ell^2(\mathbb{Z})$. In fact $\mathscr{F}_\Omega$ is a Riesz basis of $\ell^2(\mathbb{Z})$. Examples in which the Ω is nonuniform can be found in [**?**].

Proof of Theorem 2.4.6.

Again, using Remark 2.4.1, we will show that $\{D^l b : l = 0, 1, \dots\}$ is not minimal. We first assume that $D = \sum_j \lambda_j P_j$ is non-negative, i.e., $\lambda_j \geq 0$ for all $j \in \mathbb{N}$. Since $A \in B(\ell^2(\mathbb{N}))$, we also have that $0 \leq \lambda_j \leq \|D\| < \infty$. Let $b \in O_\Omega$ be such that its D-annihilator has degree $r = \infty$ and let n_k be any increasing sequence of nonnegative integers such that $\sum_{k=2}^{\infty} 1/n_k = \infty$.

Fix $f \in \overline{span}\{D^l b : l = 0, 1, \dots\}$. Then for any $\varepsilon > 0$, there exists a polynomial p such that $\|f - p(D)b\|_2 \leq \varepsilon/2$. Since the polynomial p is a continuous function on $C[0, \|D\|]$, (by the Müntz-Szász Theorem) there exists a polynomial $g \in span\{1, x^{n_k} : k \in \mathbb{N}\}$ such that $\sup\{|p(x) - g(x)| : x \in [0, \|D\|]\} \leq \frac{\varepsilon}{2\|b\|_2}$.

Now we note that

$$\|p(D)b - g(D)b\|^2_{\ell^2(\mathbb{N})} = \sum_j |p(\lambda_j) - g(\lambda_j)|^2 |b_j|^2 \leq (\varepsilon/2)^2.$$

Hence

$$\|f - g(D)b\|_2 \quad \leq \quad \|f - p(D)b\|_2 + \|p(D)b - g(D)b\|_2 \leq \varepsilon$$

Therefore $\overline{span}\{b, D^{n_k} b : k \in \mathbb{N}\} = \overline{span}\{D^l b : l = 0, 1, \dots\}$ and we conclude that $\{D^l b : l = 0, 1, \dots\}$ is not minimal.

If the assumption about the non-negativity of $D = \sum_j \lambda_j P_j$ is removed, then by the previous argument $\{D^{2l} b : l = 0, 1, \dots\}$ is not minimal hence $\{D^l b : l = 0, 1, \dots\}$ is not minimal either, and the proof is complete.

□

The following corollary of Theorem 2.4.6 will be needed later.

Corollary 2.4.10. *Let b be such that its D-annihilator has degree $r = \infty$. If there exists an increasing sequence $\{n_k : k \in \mathbb{N}\}$ of positive integers such that $\sum_{k=2}^{\infty} \frac{1}{n_k} = +\infty$, then the collection $\{D^{n_k} b : k \in \mathbb{N}\}$ is not minimal.*

Proof. Pick a subsequence $\{n_{k_j}\}$ of $\{n_k\}$ such that $\sum_{j=2}^{\infty} \frac{1}{n_{k_j}} = +\infty$ and apply the same argument as in the proof of the theorem. □

2.4.3 Frames in infinite dimensional Hilbert spaces

In the previous sections, we have seen that although the set $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ can be complete for appropriate sets Ω, it cannot form a basis for $\ell^2(\mathbb{N})$ if Ω is a finite set, in general. The main reason is that $\mathscr{F}_\Omega$ cannot be minimal, which is necessary to be a basis. On the other hand, the non-minimality is a statement about redundancy. Thus, although $\mathscr{F}_\Omega$ cannot be a basis, it is possible that $\mathscr{F}_\Omega$ is a frame for sets $\Omega \subseteq \mathbb{N}$ with finite cardinality. Being a frame is in fact desirable since in this case we can reconstruct any $f \in \ell^2(\mathbb{N})$ in stable way from the data $Y = \{f(i), Af(i), A^2 f(i), \ldots, A^{l_i} f(i) : i \in \Omega\}$.

In this section we will show that, except for some special case of the eigenvalues of A, if Ω is a finite set, i.e., $|\Omega| < \infty$, then $\mathscr{F}_\Omega$ can never be a frame for $\ell^2(\mathbb{N})$. Thus essentially, either the eigenvalues of A are *nice*, as we will make precise below, in which case we can choose Ω to consist of just one element whose iterations may be a frame, or, the only hope for $\mathscr{F}_\Omega$ to be a frame for $\ell^2(\mathbb{N})$ is that Ω is infinite in which case it needs to be well-spread over $\mathbb{N}$.

Theorem 2.4.11. *Let $A \in \mathscr{A}$ and let $\Omega \subseteq \mathbb{N}$ be a finite subset of $\mathbb{N}$. If $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ is a frame, with constants C_1 and C_2, then*

$$\inf\{\|A^l e_i\|_2 : \ i \in \Omega, l = 0, \ldots, l_i\} = 0.$$

Proof.

If $\mathscr{F}_\Omega$ is a frame, then it is complete. Therefore, since Ω is finite, there exists $i_0 \in \Omega$

with $l_{i_0} = +\infty$. We have:

$$\sum_{l=0}^{+\infty} \|A^l e_{i_0}\|^4 = \sum_{l=0}^{+\infty} |< A^l e_{i_0}, A^l e_{i_0} >|^2 = \sum_{l=0}^{+\infty} |< e_{i_0}, A^{2l} e_{i_0} >|^2 \leq C_2.$$

As a consequence $\|A^l e_{i_0}\|$ goes to zero with l. □

Therefore, when $|\Omega| < \infty$, the only possibility for $\mathscr{F}_\Omega$ to be a frame, is that

$$\inf\{\|A^l e_i\|_2 : \ i \in \Omega, l = 0, \ldots, l_i\} = 0$$

and

$$\sup\{\|A^l e_i\|_2 : \ i \in \Omega, l = 0, \ldots, l_i\} \leq C < \infty.$$

We have the following theorem to establish for which finite sets Ω, $\mathscr{F}_\Omega$ is not a frame for $\ell^2(\mathbb{N})$.

Theorem 2.4.12. *Let $A \in \mathscr{A}$ and let Ω be a finite subset of $\mathbb{N}$. For $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ to be a frame, it is necessary that 1 or -1 are cluster points of $\sigma(A)$.*

Since a compact self-adjoint operator on a Hilbert space either has finitely many eigenvalues or the eigenvalues form a sequence that goes to zero, we have the following corollary:

Corollary 2.4.13. *Let A be a compact self-adjoint operator, and $\Omega \subseteq \mathbb{N}$ be a finite set. Then $\mathscr{F}_\Omega = \left\{A^l e_i : \ i \in \Omega, l = 0, \ldots, l_i\right\}$ is not a frame.*

Remark 2.4.14. Theorems 2.4.11 and 2.4.12 can be generalized to the class $\mathscr{A}$ defined in (ii) of Remark 2.4.4.

Proof of Theorem 2.4.12.

If $\mathscr{F}_\Omega$ is a frame then it is complete in $\ell^2(\mathbb{N})$, then the set $\Omega_\infty := \{i \in \Omega : l_i = \infty\}$ is nonempty.

Using again the completeness of $\mathscr{F}_\Omega$ we see that the set

$$J = \{j \in \mathbb{N} : P_j b_i = 0, \forall i \in \Omega_\infty\},$$

must be finite. (For this note that if J is infinite then $\bigoplus_{j\in J} E_j$ is infinite dimensional and can not be generated by de finite set of vectors $\left\{D^l b_i : \ i \in \Omega \setminus \Omega_{+\infty}, l = 0, \ldots, l_i\right\}$.

If there exists $j \in \mathbb{N}$ and $i \in \Omega_\infty$ such that $|\lambda_j| \geq 1$ and $P_j b_i \neq 0$ then for $x = P_j b_i$ we have

$$\sum_l |\langle x, D^l b_i\rangle|^2 = \sum_l |\lambda_j|^{2l} \|P_j b_i\|_2^4 = \infty.$$

Thus, $\mathscr{F}_\Omega$ is not a frame.

Otherwise, let $r := \sup_{j\in\mathbb{N}}\{|\lambda_j| : P_j b_i \neq 0 \text{ for some } i \in \Omega_\infty\}$.
Since -1 or 1 are not cluster points of $\sigma(A)$, $r < 1$. But

$$\|Db_i\|_2 \leq \sup_{j\in\mathbb{N}}\{|\lambda_j| : P_j b_i \neq 0\}\|b_i\|_2 \quad \forall i \in \Omega_\infty,$$

and therefore we have that $\|D^l b_i\|_2 \leq r^l \|b_i\|_2$. Now given $\varepsilon > 0$, there exists N such that

$$\sum_{i\in\Omega_\infty}\sum_{l>N} \|D^l b_i\|_2^2 \leq \varepsilon.$$

Choose $f \in \ell^2(\mathbb{N})$ such that $\|f\|_2 = 1$, $< f, D^l b_i >= 0$ for all $i \in \Omega - \Omega_\infty$ and $l = 0, \ldots, l_i$ and such that $< f, D^l b_i >= 0$ for all $i \in \Omega_\infty$ and $l = 0, \ldots, N$. Then

$$\sum_{i\in\Omega}\sum_{l=0}^{l_i} | < f, D^l b_i > |^2 \leq \varepsilon = \varepsilon \|f\|_2.$$

Since ε is arbitrary, the last inequality implies that $\mathscr{F}_\Omega$ is not a frame since it cannot have a positive lower frame bound. □

Although Theorem 2.4.12 states that $\mathscr{F}_\Omega$ is not a frame for $\ell^2(\mathbb{N})$, it could be that after normalization of the vectors in $\mathscr{F}_\Omega$, the new set $\mathscr{Z}_\Omega$ is a frame for $\ell^2(\mathbb{N})$. It turns out that

the obstruction is intrinsic. In fact, this case is even worse, since $\mathscr{Z}_\Omega$ is not a frame even if 1 or -1 is (are) a cluster point(s) of $\sigma(A)$.

Theorem 2.4.15. *Let $A \in \mathscr{A}$ and let $\Omega \subseteq \mathbb{N}$ be a finite set. Then the unit norm sequence $\left\{\frac{A^l e_i}{\|A^l e_i\|_2} : i \in \Omega, l = 0,\dots,l_i\right\}$ is not a frame.*

Proof. Note that by Remark 2.4.1, $\left\{\frac{A^l e_i}{\|A^l e_i\|_2} : i \in \Omega, l = 0,\dots,l_i\right\}$ is a frame if and only if $\mathscr{Z}_\Omega = \left\{\frac{D^l b_i}{\|D^l b_i\|_2} : i \in \Omega, l = 0,\dots,l_i\right\}$ is a frame.

Assume that $\mathscr{Z}_\Omega$ is a frame. Since it is norm-bounded (actually unit norm), the Kadison-Singer/Feichtinger conjectures proved recently [37] applies, and $\mathscr{Z}_\Omega$ is the finite union of Riesz sequences $\bigcup_{j=1}^{N} R_j$.

Because $\mathscr{Z}_\Omega$ is complete, there exist some b such that its D-annihilator has degree $r = \infty$, $j \in \{1,\dots,N\}$ and an increasing sequence of positive integers $\{n_k\}$ with $\sum_{k\geq 2} \frac{1}{n_k} = +\infty$ such that

$$S = \left\{ \frac{D^{n_k} b}{\|D^{n_k} b\|_2} : k \in \mathbb{N} \right\} \subseteq R_j.$$

The set S is a Riesz sequence, because it is a subset of a Riesz sequence. On the other hand, S is not minimal by Corollary 2.4.10, which is a contradiction since a Riesz sequence is always a minimal set.

□

We will now concentrate on the case when there is a cluster point of $\sigma(A)$ at 1 or -1, and we start with the case where Ω consists of a single sampling point, i.e., $O_\Omega = \{b\}$. Let us denote by r_b, the degree of the D-annihilator of b and $l_b = r_b - 1$ if r_b is finite or $l_b = +\infty$ otherwise.

Since $A \in \mathscr{A}$, $A = B^* D B$, by Remark 2.4.1 $\mathscr{F}_\Omega$ is a frame of $\ell^2(\mathbb{N})$ if and only if there exists a vector $b = Be_j$ for some $j \in \mathbb{N}$ that corresponds to the sampling point, and $\{D^l b : l = 0,1,\dots\}$ is a frame for $\ell^2(\mathbb{N})$.

For this case, Theorem 2.4.3 implies that if $\mathscr{F}_\Omega$ is a frame of $\ell^2(\mathbb{N})$, then the projection operators P_j used in the description of the operator $A \in \mathscr{A}$ must be of rank 1. Moreover,

the vector b corresponding to the sampling point must have infinite support, otherwise l_b will be finite and $\mathscr{F}_\Omega$ cannot be complete in $\ell^2(\mathbb{N})$. Moreover, for this case in order for $\mathscr{F}_\Omega$ to be a frame, it is necessary that $|\lambda_k| < 1$ for all k, otherwise, if there exists $\lambda_{j_0} \geq 1$ then for $x = P_{j_0} b$ (note that by Theorem 2.4.3 $P_{j_0} b \neq 0$) we would have

$$\sum_n |\langle x, D^n b\rangle|^2 = \sum_n |\lambda_{j_0}|^{2n} \|P_{j_0} b\|_2^4 = \infty,$$

which is a contradiction.

In addition, if $\mathscr{F}_\Omega$ is a frame, then the sequence $\{\lambda_k\}$ cannot have a cluster point a with $|a| < 1$. To see this, suppose there is a subsequence $\lambda_{k_s} \to a$ for some a with $|a| < 1$, and let W be the ortogonal sum of the eigenspaces associated to the eigenvalues λ_{k_s}. Then W is invariant for D. Set $D_1 = D|_W$, and $\tilde{b} = P_W b$ where P_W is the orthogonal projection on W. Then, by Theorem 2.4.12, $\{D_1^j \tilde{b} : j = 0, 1, \dots\}$ can not be a frame for W. It follows that $\mathscr{F}_\Omega$ cannot be a frame for $\ell^2(\mathbb{N})$, since the ortogonal projection of a frame onto a closed subspace is a frame of the subspace.

Thus the only possibility for $\mathscr{F}_\Omega$ to be a frame of $\ell^2(\mathbb{N})$ is that $|\lambda_k| \to 1$. These remarks allow us to characterize when $\mathscr{F}_\Omega$ is a frame for the situation when $|\Omega| = 1$.

Theorem 2.4.16. *Let $D = \sum_j \lambda_j P_j$ be such that P_j have rank 1 for all $j \in \mathbb{N}$, and let $b := \{b(k)\}_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Then $\{D^l b : l = 0, 1, \dots\}$ is a frame if and only if*

i) $|\lambda_k| < 1$ for all k.

ii) $|\lambda_k| \to 1$.

iii) $\{\lambda_k\}$ satisfies Carleson's condition

$$\inf_n \prod_{k\neq n} \frac{|\lambda_n - \lambda_k|}{|1 - \bar{\lambda}_n \lambda_k|} \geq \delta. \tag{2.17}$$

for some $\delta > 0$.

iv) $b(k) = m_k\sqrt{1-|\lambda_k|^2}$ *for some sequence* $\{m_k\}$ *satisfying* $0 < C_1 \leq |m_k| \leq C_2 < \infty$.

This theorem implies the following Corollary:

Corollary 2.4.17. *Let* $A = B^*DB \in \mathscr{A}$, *and* $D = \sum_j \lambda_j P_j$ *be such that* P_j *have rank* 1 *for all* $j \in \mathbb{N}$. *Then, there exists* $i_0 \in \mathbb{N}$ *such that* $\mathscr{F}_\Omega = \{A^l e_{i_0} : l = 0, \dots\}$ *is a frame for* $\ell^2(\mathbb{N})$, *if and only if* $\{\lambda_j\}$ *satisfy the conditions of Theorem 2.4.16 and there exists* $i_0 \in \mathbb{N}$, *such that* $b = Be_{i_0}$ *satisfies the condition* iv *of Theorem 2.4.16* .

Theorem 2.4.16 follows from the discussion above and the following two Lemmas

Lemma 2.4.18. *Let D be as in Theorem 2.4.16 and assume that* $|\lambda_k| < 1$ *for all k. Let* $b^0(k) = \sqrt{1-|\lambda_k|^2}$, *and assume that* $b^0 \in \ell^2(\mathbb{N})$. *Let* $b \in \ell^2(\mathbb{N})$.

Then, $\{D^l b : l \in \mathbb{N}\}$ *is a frame for* $\ell^2(N)$ *if and only if* $\{D^l b^0 : l \in \mathbb{N}\}$ *is a frame and there exist* C_1 *and* C_2 *such that* $b(k)/b^0(k) = m_k$ *satisfies* $0 < C_1 \leq |m_k| \leq C_2 < \infty$.

Note that by assumption $\sum_{k=1}^{\infty}(1-|\lambda_k|^2) < +\infty$ since $b^0 \in \ell^2(\mathbb{N})$. In particular $|\lambda_k| \to 1$.

Lemma 2.4.19. *Let* $D = \sum_j \lambda_j P_j$ *be such that* $|\lambda_k| < 1$, $\lambda_k \longrightarrow 1$ *and let* $b^0(k) = \sqrt{1-|\lambda_k|^2}$. *Then the following are equivalent:*

i)

$$\{b^0, Db^0, D^2b^0, \dots\} \text{ is a frame for } \ell^2(N)$$

ii)

$$\inf_n \prod_{k\neq n} \frac{|\lambda_n - \lambda_k|}{|1-\bar{\lambda}_n\lambda_k|} \geq \delta.$$

for some $\delta > 0$.

In Lemma 2.4.19, the assumption $\lambda_k \longrightarrow 1$ can be replaced by $\lambda_k \longrightarrow -1$ and the lemma remains true. Its proof, below, is due to J. Antezana [8] and is a consequence of a theorem by Carleson [28] about interpolating sequences in the Hardy space $H^2(\mathbb{D})$ of the unit disk in $\mathbb{C}$.

Proof of Lemma 2.4.18.

Let us first prove the sufficiency. Assume that $\{D^l b^0 : l \in \mathbb{N}\}$ is a frame for $\ell^2(\mathbb{N})$ with positive frame bounds A, B, and let $b \in \ell^2(\mathbb{N})$ such that $b(k) = m_k b^0(k)$ with $0 < C_1 \leq |m_k| \leq C_2 < \infty$. Let $x \in \ell^2(\mathbb{N})$ be an arbitrary vector and define $y_k = \overline{m_k} x_k$. Then $y \in \ell^2(\mathbb{N})$ and $C_1\|x\|_2 \leq \|y\| \leq C_2\|x\|_2$. Hence

$$C_1^2 A\|x\|_2^2 \leq \sum_l |\langle y, D^l b^0\rangle|^2 = \sum_l \langle x, D^l b\rangle|^2 \leq C_2^2 B\|x\|_2^2,$$

and therefore $\{D^l b : l \in \mathbb{N}\}$ is a frame for $\ell^2(\mathbb{N})$.

Conversely, let $b \in \ell^2(\mathbb{N})$ and assume that $\{D^l b : l \in \mathbb{N}\}$ is a frame for $\ell^2(N)$ with frame bounds A' and B'. Then for any vector e_k of the standard orthonormal basis of $\ell^2(\mathbb{N})$, we have

$$A' \leq \sum_{l=0}^{\infty} | < e_k, D^l b > |^2 = \frac{|b(k)|^2}{1 - |\lambda_k|^2} \leq B'.$$

Thus $\sqrt{A'} b^0(k) \leq |b(k)| \leq \sqrt{B'} b^0(k)$ for all k. Thus, the sequence $\{m_k\} \subseteq \mathbb{C}$ defined by $b(k) = m_k b^0(k)$ satisfies $\sqrt{A'} \leq |m_k| \leq \sqrt{B'}$.

Let $x \in \ell^2(\mathbb{N})$ be an arbitrary vector and define now $y_k = \frac{1}{\overline{m_k}} x_k$. Then $y \in \ell^2(\mathbb{N})$ and

$$\frac{A'}{B'}\|x\|_2^2 \leq \sum_l |\langle x, D^l b^0\rangle|^2 = \sum_l |\langle y, D^l b\rangle|^2 \leq \frac{B'}{A'}\|x\|_2^2.$$

and so $\{D^l b^0 : l \in \mathbb{N}\}$ is a frame for $\ell^2(\mathbb{N})$. □

The proof of Lemma 2.4.19 relies on a Theorem by Carleson on interpolating sequences in the Hardy space $H^2(\mathbb{D})$ on the open unit disk $\mathbb{D}$ in the complex plane. If $H(\mathbb{D})$ is the vector space of holomorphic functions on $\mathbb{D}$, $H^2(\mathbb{D})$ is defined as

$$H^2(\mathbb{D}) = \Big\{ f \in H(\mathbb{D}) : f(z) = \sum_{n=0}^{\infty} a_n z^n \text{ for some sequence } \{a_n\} \in \ell^2(\mathbb{N}) \Big\}.$$

Endowed with the inner product between $f = \sum_{n=0}^{\infty} a_n z^n$ and $g = \sum_{n=0}^{\infty} a'_n z^n$ defined by

$\langle f,g\rangle = \sum a_n\overline{a'_n}$, $H^2(\mathbb{D})$ becomes a Hilbert space isometrically isomorphic to $\ell^2(\mathbb{N})$ via the isomorphism $\Phi(f) = \{a_n\}$.

Definition 2.4.20. A sequence $\{\lambda_k\}$ in $\mathbb{D}$ is an interpolating sequence for $H^2(\mathbb{D})$ if for any sequence $\{c_k\}$ such that $\sum_k |c_k|^2(1-|\lambda_k|^2) < +\infty$, there exists a function $f \in H^2(\mathbb{D})$ such that $f(\lambda_k) = c_k$.

Proof of Lemma 2.4.19 .

Let $\mathscr{T}_k$, denote the vector in $\ell^2(\mathbb{N})$ defined by $\mathscr{T}_k = (1, \lambda_k, \lambda_k^2, \dots)$, and $x \in \ell^2(\mathbb{N})$. Then

$$\sum_{l=0}^{\infty} |< x, D^l b^0 >|^2 = \sum_{l=0}^{\infty} \Big| \sum_{k=1}^{\infty} x_k \lambda_k^l \sqrt{1-\lambda_k^2}\Big|^2 = \sum_{s=1}^{\infty}\sum_{t=1}^{\infty} \frac{< \mathscr{T}_s, \mathscr{T}_t >}{\|\mathscr{T}_s\|_2 \|\mathscr{T}_t\|_2} x_s \overline{x_t}.$$

Thus, for $\{D^l b^0 : l = 0, 1, \dots\}$ to be a frame of $\ell^2(\mathbb{N})$, it is necessary and sufficient that the Gramian $G_\Lambda = \{G_\Lambda(s,t)\} = \left\{ \frac{< \mathscr{T}_s, \mathscr{T}_t >}{\|\mathscr{T}_s\|_2 \|\mathscr{T}_t\|_2} \right\}$ be a bounded invertible operator on $\ell^2(N)$ (Note that G_Λ is then the frame operator for $\{D^l b^0 : l = 0, 1, \dots\}$).

Equivalently, $\{D^l b^0 : l = 0, 1, \dots\}$ is a frame of $\ell^2(\mathbb{N})$ if and only if the sequence $\{\widetilde{\mathscr{T}_j} = \frac{\mathscr{T}_j}{\|\mathscr{T}_j\|_2}\}$ is a Riesz basic sequence in $\ell^2(\mathbb{N})$, i.e., there exist constants $0 < C_1 \le C_2 < \infty$ such that

$$C_1 \|c\|_2^2 \le \| \sum_j c_j \widetilde{\mathscr{T}_j} \|_2^2 \le C_2 \|c\|_2^2 \quad \text{for all } c \in \ell^2(\mathbb{N}).$$

By the isometric map Φ from $\ell^2(\mathbb{N})$ to $H^2(\mathbb{D})$ defined above, $\{D^l b^0 : l = 0, 1, \dots\}$ is a frame of $\ell^2(\mathbb{N})$ is a frame if and only if the sequence $\{\tilde{k}_{\lambda_j} = \Phi(\widetilde{\mathscr{T}_j})$ is a Riesz basic sequence in $H^2(\mathbb{D})$.

Let $k_{\lambda_j} = \Phi(\mathscr{T}_j)$. It is not difficult to check that for any $f \in H^2(\mathbb{D})$, $\langle f, k_{\lambda_j}\rangle = f(\lambda_j)$ and that $\{\lambda_j\}$ is an interpolating sequence in $H^2(\mathbb{D})$ if and only if $G_\Lambda = \left(\langle \tilde{k}_{\lambda_j}, \tilde{k}_{\lambda_j}\rangle\right)$ is a bounded invertible operator on $\ell^2(\mathbb{N})$. By Carleson' s Theorem [28], this happens if and only if (2.17) is satisfied. □

Frames of the form $\{D^l b_i : i \in \Omega, l = 0 \dots, l_i\}$ for the case when $|\Omega| \ge 1$ or when the projections P_j have finite rank but possibly greater than or equal to 1 can be easily found by

using Theorem 2.4.16. For example, if $|\Omega| = 2$, $P_j(\ell^2(\mathbb{N}))$ has dimension 1 for $j \in \mathbb{N}$, b_1, $\{\lambda_k\}$ satisfies the conditions of Theorem 2.4.16 and b_2 is such that $b_2(k) = m_k\sqrt{1-|\lambda_k|^2}$ for some sequence $\{m_k\}$ satisfying $|m_k| \leq C < \infty$. To construct frames for the case when the projections P_j have finite rank but possibly greater than or equal to 1, we note that there exist orthogonal subspaces $W_1, \ldots, W_N$ of $\ell^2(\mathbb{N})$ such that operator D_i on each W_i either has finite dimensional range, or satisfies the condition of Theorem 2.4.16.

2.5 Concluding Remarks

In this chapter we have studied the sets of spatial sampling locations Ω that allow us to reconstruct a function f from the samples of $\{f(i), Af(i), \ldots, A^{l_i}f(i) : i \in \Omega\}$. The finite dimensional case is completely resolved and we find necessary and sufficient conditions on Ω, l_i, and A for the stable recovery of f.

For the case where $\mathscr{H} \approx \ell^2(\mathbb{N})$, we restricted ourselves to the subclass $\mathscr{A}$ of self-adjoint diagonalizable operators. Without stability requirements, the sets Ω for which a reconstruction of f is possible are completely characterized. For the case where Ω is an infinite set, there are examples for which the stable reconstruction of f is possible as in [6], and it is not difficult to construct other examples of infinite sets Ω for which stable reconstruction is possible as well. However, the problem of finding necessary and sufficient conditions for stable reconstruction is still open.

Chapter 3

SYSTEM IDENTIFICATION IN DYNAMICAL SAMPLING

3.1 Problem Formulation

In this section, we formulate a special instance of system identification problem in the 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 the spatiotemporal samples with both x and A unknown, we would like to recover as much information about them as we can under the given various priors. Here we first study the case of uniform subsampling. Without loss of generality, let a positive odd integer m ($m > 1$) be the uniform subsampling factor. At time level $t = l$, we uniformly undersample the evolving state $A^l x$ and get the spatiotemporal data

$$y_l = (a^l * x)(m\mathbb{Z}), \tag{3.1}$$

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

Problem 3.1.1. 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 $\{x(m\mathbb{Z}), (a * x)(m\mathbb{Z}), \cdots, (a^{N-1} * x)(m\mathbb{Z})\}$?

3.1.1 Previous work

In [7], 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 $\Omega \subseteq \mathbb{Z}_d = \{1, 2, \cdots, d\}$, 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 regular invariant dynamical sampling problem in [7]. 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 [6], 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 [40], 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 it, but deals with a fundamentally different problem. In Sparse Fourier Transformation, see [22, 33, 34], the idea is to uniformly undersample the 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 of the shape of an evolving wave by observing the amplitude of its aliasing 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 include the the Slepian-Wolf distributed source coding problem [42] and the distributed sampling problem in [30]. Our problem, however, is very different from the above in the nature of the processes we study. 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.

3.1.2 Notations and terminologies

In the following contents of chapter 3, 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 the known submatrix notation coincides with 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 3.1.2. The minimal annihilating polynomial of a square matrix $\mathbf{A}_M$ is $p^{\mathbf{A}_M}[z]$, if it 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 $deg(p^{\mathbf{A}_M})$.

Let the 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 3.1.3. Let $w_1, w_2, \cdots, w_n$ be n distinct complex numbers, denote $\mathbf{w} = [w_1, \cdots, w_n]^T$, the $n \times N$ Vandermonde matrix generated by $\mathbf{w}$ is defined by

$$\mathbf{V}_{n,N}(\mathbf{w}) = \begin{pmatrix} 1 & w_1 & \cdots & w_1^{N-1} \\ 1 & w_2 & \cdots & w_2^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w_n & \cdots & w_n^{N-1} \end{pmatrix}. \tag{3.2}$$

Definition 3.1.4. For a sequence $c = (c_n)_{n\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$ or $\ell^2(\mathbb{Z})$, we define its Fourier transformation to be the function on the Torus $\mathbb{T} = [0,1)$

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

3.2 Contribution and Organization

The remainder of the this chapter is organized as follows: In section 3.3, we discuss the noise free case. From a theoretical aspect, we 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 3.4, 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 3.5, we do several numerical experiments to verify some estimation results. In section 3.6, we propose several other algorithms such as Generalized Matrix Pencil method, Generalized ESPRIT Method and Cadzow Denoising methods to improve the effectiveness and robustness of recovery. The comparison between algorithms is illustrated by a numerical example in section 3.7. Finally, we summarize the work in section 3.8.

3.3 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" so that $\hat{a}(\xi)$ is real, symmetric and strictly decreasing on $[0, \frac{1}{2}]$. An example of such a typical low pass filter is shown in Figure 1. 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

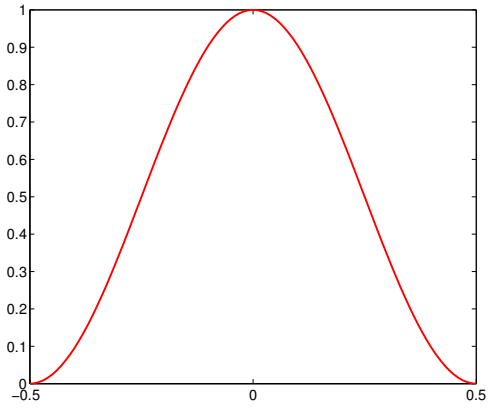

Figure 3.1: A Typical Low Pass Filter

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 exact. 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 3.3.1. *The Fourier transform of each measurement sequence $y_l = S_m(a^l * x)$ at $\xi \in \mathbb{T}$ is*

$$\hat{y}_l(\xi) = \frac{1}{m}\sum_{i=0}^{m-1} \hat{a}^l(\frac{\xi+i}{m})\hat{x}(\frac{\xi+i}{m}). \tag{3.3}$$

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

$$\mathbf{h}_t(\xi) = [\hat{y}_t(\xi), \hat{y}_{t+1}(\xi), \cdots, \hat{y}_{N-m+t-1}(\xi)]^T, \tag{3.4}$$

and build the Hankel matrices

$$\mathbf{H}^{\xi}_{N-m,m}(0) = \big[\mathbf{h}_0(\xi), \mathbf{h}_1(\xi), \cdots, \mathbf{h}_{m-1}(\xi)\big], \tag{3.5}$$

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

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

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

$$m\mathbf{H}^{\xi}_{N-m,m}(s) = \mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi))diag(\mathbf{x}(\xi))diag(\mathbf{w}(\xi))^s\mathbf{V}_m(\mathbf{w}(\xi)), \tag{3.6}$$

   where $s = 0, 1$. The Vandermonde matrix $\mathbf{V}_{m,N-m}(\mathbf{w}(\xi))$ and $\mathbf{V}_m(\mathbf{w}(\xi))$ are given in the way as indicated in Definition 3.2.

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:

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

3. *Assume the entries of* $\mathbf{x}(\xi)$ *are all nonzero. For* $\xi \neq 0, \frac{1}{2}$*, the vector defined by*

$$\mathbf{q}(\xi) = [q_0(\xi), \cdots, q_{m-1}(\xi)]^T$$

is the unique solution of the linear system

$$\mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{q}(\xi) = -\mathbf{h}_m(\xi) \tag{3.7}$$

if and only if the polynomial

$$q^{\xi}[z] = \sum_{k=0}^{m-1} q_k(\xi) z^k + z^m \tag{3.8}$$

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*

$$\mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{C}^{p[z]} = \mathbf{H}^{\xi}_{N-m,m}(1) \tag{3.9}$$

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

Proof. (1) By Lemma 3.3.1, 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 $\mathbf{H}^{\xi}_{N-m,m}(0)$. Notice 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

$$\text{Rank } V_m(\mathbf{w}(\xi)) = \begin{cases} m & \text{if } \xi \neq 0 \text{ or } \frac{1}{2}, \\ \frac{m+1}{2} & \text{otherwise.} \end{cases} \tag{3.10}$$

Since $N \geq 2m$, Rank $\mathbf{V}_m(\mathbf{w}(\xi)) =$ 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 results in (1).

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

$$\begin{aligned} \mathbf{H}^{\xi}_{N-m,m}(0)\mathbf{q}(\xi) &= \sum_{k=0}^{m-1} q_k(\xi)\mathbf{h}_k(\xi) \\ &= \mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi))(\sum_{k=0}^{m-1} q_k(\xi)diag(\mathbf{w}(\xi))^k)\mathbf{x}(\xi) \\ &= -\mathbf{V}^T_{m,N-m}(\mathbf{w}(\xi))diag(\mathbf{w}(\xi))^m\mathbf{x}(\xi) \\ &= -\mathbf{h}_m(\xi). \end{aligned} \tag{3.11}$$

Conversely, if $\mathbf{q}(\xi)$ is the solution of linear system (3.7), let the monic polynomial given by $\mathbf{q}(\xi)$ be $q^{\xi}[z]$, then by the computation process of (3.11), 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, $q^{\xi}[diag(\mathbf{w}(\xi))]\mathbf{x}(\xi) = 0$. Since $q^{\xi}[diag(\mathbf{w}(\xi))]$ is diagonal and $\mathbf{x}(\xi)$ has no zero entries, we know $q^{\xi}[z]$ is a monic annihilating polynomial of $diag(\mathbf{w}(\xi))$. The minimality is followed by counting its degree. If $p[z]$ is a monic annihilating polynomial of $diag(\mathbf{w}(\xi))$, by computations, it is easy to show the identity (3.9) is an equivalent formulation with the identity (3.7). □

Corollary 3.3.3. *In the case of $\xi = 0$ 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:*

$$\mathbf{H}^{\xi}_{N-m,\frac{m+1}{2}}(0)\mathbf{c}(\xi) = -\mathbf{h}_{\frac{m+1}{2}}(\xi), \tag{3.12}$$

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 3.3.4. *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* (3.1) *when $N \geq 2m$.*

Proof. Since Fourier transformation 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 of 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 $E = E_0 - \{0, \frac{1}{2}\}$, if $\xi \in E$, by (3) of Proposition 3.3.2, we can recover the minimal annihilating polynomial of $diag(\mathbf{w}(\xi))$. 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 symmetric 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 transformation 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 [6]. □

Theorem 3.3.4 addresses the infinite dimensional analog of Theorem 4.1 in [7]. 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 $\mathbb{T}$, where the measure of this subset is 1.

Definition 3.3.5. 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 3.3.4, 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 $\mathbb{T}$. It is summarized in Algorithm 3.3.1.

Algorithm 3.3.1 Generalized Prony Method

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

1: Compute the Fourier transformation 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 (3.7):

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

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 symmetric condition of $\hat{a}$ to get $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

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

Corollary 3.3.6. *In addition to the assumptions of Theorem 3.3.4, 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 3.3.1.*

Proof. Under these assumptions, we know

$$\hat{a}(\xi) = a(0) + \sum_{k=1}^{r} a(k)\cos(2\pi k\xi). \tag{3.13}$$

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

$$\begin{pmatrix} 1 & cos(2\pi\eta_1) & \cdots & cos(2r\pi\eta_1) \\ 1 & cos(2\pi\eta_2) & \cdots & cos(2r\pi\eta_2) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & cos(2\pi\eta_r) & \cdots & cos(2r\pi\eta_r) \end{pmatrix} \begin{pmatrix} a(0) \\ a(1) \\ \vdots \\ a(r) \end{pmatrix} = \begin{pmatrix} \hat{a}(\eta_1) \\ \hat{a}(\eta_2) \\ \vdots \\ \hat{a}(\eta_r) \end{pmatrix}. \tag{3.14}$$

Note that $\{1, \cos(2\pi\eta), \cdots, \cos(2r\pi\eta)\}$ is a Chebyshev system on $[0,1]$(see [46]), and

hence (3.14) 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 ξs, which finishes the proof. □

3.4 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 3.3.1. 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 3.3.1 in the presence of noise. Mathematically speaking, assume the measurements are given by $\{\tilde{y}_l\}_{l=0}^{N-1}$ compared to (3.1) so that $||\hat{y}_l(\xi) - \widehat{\tilde{y}_l}(\xi)||_\infty \leq \varepsilon_l$ for all $\xi \in \mathbb{T}$. Given an estimation for $\varepsilon = max_l|\varepsilon_l|$, how large can the error be in the worst case for the output parameters of Algorithm 3.3.1 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 the Algorithm 3.3.1.

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 firstly measure the perturbation of $\mathbf{q}(\xi)$ in terms of ℓ^∞ norm. Secondly we measure the perturbation of the roots. It is well known that the roots of a polynomial are continuously dependent on the small change of its coefficients, see Proposition 3.4.2. 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 3.4.1. Let $\xi \in \mathbb{T} - \{0, \frac{1}{2}\}$, consider the set $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ that consists of m distinct nodes.

1. For $0 \leq k \leq m-1$, the separation between $\hat{a}(\frac{\xi+k}{m})$ with 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
$$\sigma_k(\xi) = \begin{cases} 1 & \text{if } k\text{=0}, \\ \sum\limits_{0 \leq j_1 < \cdots < j_k \leq m-1} \hat{a}(\frac{\xi+j_1}{m})\hat{a}(\frac{\xi+j_2}{m}) \cdots \hat{a}(\frac{\xi+j_k}{m}) & \text{otherwise.} \end{cases} \tag{3.15}$$
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 3.4.2 (see Proposition **V**.1 in [9])**.** *Let z_k be a root of multiplicity $M_k \in \mathbb{N}^+$ of the r-th order polynomial $p[z]$. For all $\varepsilon > 0$, let $p_\varepsilon[z] = p[z] + \varepsilon\Delta p[z]$, 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 $\varepsilon < \varepsilon_0$ there are exactly M_k roots of $p_\varepsilon[z]$, denoted $\{z_{k,m}(\varepsilon)\}_{m \in \{0,\cdots,M_k-1\}}$, which admit the first-order fractional expansion*

$$z_{k,m}(\varepsilon) = z_k + \varepsilon^{\frac{1}{M_k}} \Delta z_k e^{2\pi i \frac{m}{M_k}} + O(\varepsilon^{\frac{2}{M_k}}), \tag{3.16}$$

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

$$(\Delta z_k)^{M_k} = -\frac{\Delta p[z_k]}{\frac{1}{M_k!}p^{(M_k)}[z_k]}. \tag{3.17}$$

Proposition 3.4.3. *Let the perturbed measurements* $\{\tilde{y}_l\}_{l=0}^{2m-1}$ *be given with an error satisfying* $||\widehat{\tilde{y}}_l(\xi) - \hat{y}_l(\xi)||_\infty \leq \varepsilon, \forall l$. *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* (3.5) *and* (3.3.1). *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* $\varepsilon \to 0$,

$$||\mathbf{q}(\xi) - \tilde{\mathbf{q}}(\xi)||_\infty \leq ||\mathbf{H}_m^{-1}(\xi)||_\infty(1 + m\beta_1(\xi))\varepsilon + O(\varepsilon^2), \tag{3.18}$$

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

$$|\tilde{\hat{a}}(\frac{\xi+i}{m}) - \hat{a}(\frac{\xi+i}{m})| \leq C_i(\xi)(1 + m\beta_1(\xi))||\mathbf{H}_m^{-1}(\xi)||_\infty\varepsilon + O(\varepsilon^2), \tag{3.19}$$

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

Proof. Note that linear system (3.7) is perturbed to be

$$\widetilde{\mathbf{H}}_m(\xi)\tilde{\boldsymbol{q}}(\xi) = -\tilde{\mathbf{b}}_m(\xi). \tag{3.20}$$

By our assumptions, we have

$$||\Delta\mathbf{H}_m(\xi)||_\infty = ||\widetilde{\mathbf{H}}_m(\xi) - \mathbf{H}_m(\xi)||_\infty \leq m\varepsilon, \tag{3.21}$$

$$||\Delta\mathbf{b}_m(\xi)||_\infty = ||\tilde{\mathbf{b}}_m(\xi) - \mathbf{b}_m(\xi)||_\infty \leq \varepsilon. \tag{3.22}$$

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

$$\Delta\mathbf{q}(\xi) = \mathbf{H}_m^{-1}(\xi)(I + \mathbf{H}_m^{-1}(\xi)\Delta\mathbf{H}_m(\xi))^{-1}(-\Delta\mathbf{b}_m(\xi) - \Delta\mathbf{H}_m(\xi)\mathbf{q}(\xi)). \tag{3.23}$$

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

$$\Delta\mathbf{q}(\xi) = \mathbf{H}_m^{-1}(\xi)(-\Delta\mathbf{b}_m(\xi) - \Delta\mathbf{H}_m(\xi)\mathbf{q}(\xi)) + O(\varepsilon^2). \tag{3.24}$$

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

$$||\Delta\mathbf{q}(\xi)||_\infty \leq ||\mathbf{H}_m^{-1}(\xi)||_\infty(1 + m||\mathbf{q}(\xi)||_\infty)\varepsilon + O(\varepsilon^2). \tag{3.25}$$

Since $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ are the roots of $q^\xi[z]$, using Vieta's Formulas(see [47]), 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 3.4.2, and denote by $(q^\xi)'[z]$ the derivative function of $q^\xi[z]$, for $0 \leq i \leq m-1$, we conclude

$$\begin{aligned}
|\tilde{\hat{a}}(\frac{\xi+i}{m}) - \hat{a}(\frac{\xi+i}{m})| &= |\frac{\Delta\mathbf{q}(\xi)[\hat{a}(\frac{\xi+i}{m})]}{(q^\xi)'[\hat{a}(\frac{\xi+i}{m})]} + O(\varepsilon^2)| \\
&\leq \frac{||\Delta\mathbf{q}(\xi)||_\infty(\sum_{k=0}^{m-1}|\hat{a}^k(\frac{\xi+i}{m})|)}{\prod_{\substack{j\neq i \\ 0\leq j\leq m-1}}|\hat{a}(\frac{\xi+j}{m}) - \hat{a}(\frac{\xi+i}{m})|} + O(\varepsilon^2) \\
&\leq C_i(\xi)||\mathbf{H}_m^{-1}(\xi)||_\infty(1 + m\max_{1\leq k\leq m}|\sigma_k(\xi)|)\varepsilon + O(\varepsilon^2),
\end{aligned} \tag{3.26}$$

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, i.e, a, m and x. Next, we are going to estimate $||\mathbf{H}_m^{-1}(\xi)||_\infty$ and reveal their connection with the spectral properties of a,x and the subsampling factor m.

Proposition 3.4.4. *Assume $\mathbf{H}_m(\xi)$ is invertible, we have the lower bound estimation*

$$||\mathbf{H}_m^{-1}(\xi)||_\infty \geq m \cdot \max_{i=0,\cdots,m-1} \frac{\beta_2(i,\xi)\delta_i(\xi)}{|\hat{x}(\frac{\xi+i}{m})|}, \tag{3.27}$$

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

$$||\mathbf{H}_m^{-1}(\xi)||_\infty \leq m \cdot \max_{i=0,\cdots,m-1} \frac{(\delta_i(\xi) \prod_{\substack{j\neq i \\ 0\leq j\leq m-1}} (1+|\hat{a}(\frac{\xi+j}{m})|))^2}{|\hat{x}(\frac{\xi+i}{m})|}. \tag{3.28}$$

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 the abbreviated $\mathbf{V}_m(\xi)$. Suppose $\mathbf{V}_m^{-1}(\xi) = (v_{ki})_{1\leq k,i\leq m}$ is the inverse of $\mathbf{V}_m(\xi)$, 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 $||w_i||_\infty = 1$.

$$\begin{aligned} ||\mathbf{H}_m^{-1}(\xi)||_\infty &\geq \max_{i=1,\cdots,m} ||\mathbf{H}_m^{-1}(\xi)w_i(\xi)||_\infty \\ &\geq m \cdot \max_{i=1,\cdots,m} \frac{||\mathbf{V}_m^{-1}(\xi)e_i||_\infty}{|\hat{x}(\frac{\xi+i}{m})|} \\ &= m \cdot \max_{i=0,\cdots,m-1} \frac{\beta_2(i,\xi)\delta_i(\xi)}{|\hat{x}(\frac{\xi+i}{m})|}. \end{aligned} \tag{3.29}$$

On the other hand, using the factorization (3.6) and the upper bound norm estimation for

the inverse of a Vandermonde matrix in [21], we show that

$$\begin{aligned} ||\mathbf{H}_m^{-1}(\xi)||_\infty &\leq m||\mathbf{V}_m^{-1}(\xi)||_\infty||((\mathbf{V}_m^{-1})^T(\xi))||_\infty||diag^{-1}(\mathbf{x}(\xi))||_\infty \\ &\leq m \max_{i=0,\cdots,m-1} \frac{(\delta_i(\xi) \prod_{\substack{j\neq i \\ 0\leq j\leq m-1}} (1+|\hat{a}(\frac{\xi+j}{m})|))^2}{|\hat{x}(\frac{\xi+i}{m})|}. \end{aligned} \tag{3.30}$$

□

As an application of Proposition 3.4.4, the following theorem sheds some light on the dependence of $||\mathbf{H}_m^{-1}(\xi)||_\infty$ on m.

Theorem 3.4.5. *If $|\hat{x}(\xi)| \leq M$ for every $\xi \in \mathbb{T}$, then $||\mathbf{H}_m^{-1}(\xi)||_\infty \geq O(2^m)$. Therefore, $||\mathbf{H}_m^{-1}(\xi)||_\infty \to \infty$ as $m \to \infty$.*

Proof. We show this by proving $m \cdot \max_{i=0,\cdots,m-1} \delta_i(\xi) \geq O(2^m)$. Note $\beta_2(i,x) \geq |\sigma_0^{(i)}(\xi)| = 1$. By (3.29),

$$||\mathbf{H}_m^{-1}(\xi)||_\infty \geq m \cdot \frac{\max_{i=0,\cdots,m-1} \delta_i(\xi)}{M} = O(2^m), \tag{3.31}$$

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

$$\begin{aligned} \frac{1}{c(\xi)^m} \leq \prod_{i=0}^{m-1} \frac{1}{\delta_i(\xi)} &= \prod_{0\leq i<j\leq m-1} |\hat{a}(\frac{\xi+i}{m}) - \hat{a}(\frac{\xi+j}{m})|^2 \\ &= |det(\mathbf{V}_m(\xi)|^2. \end{aligned} \tag{3.32}$$

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 [**?**]. Therefore, by the formula for the determinant of a Vandermonde matrix on the Chebyshev points in [19], we get

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

By (3.32),

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

which implies that $m \cdot c(\xi) \geq O(2^m)$. Hence by (3.31)

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

□

Remark 3.4.6. By our proof, we also see that $||\mathbf{H}_m^{-1}(\xi)||_\infty$ grows at least geometrically when m increases.

Summarizing, our results in this section suggest that

1. For $0 \leq k \leq m-1$, the accuracy of recovering the node $\hat{a}(\frac{\xi+k}{m})$ not only depends on its separation with other nodes $\delta_k(\xi)$(see Definition 3.4.1), but also depends on the global minimal separation $\delta(\xi) = \max_{k=0,\cdots,m-1} \delta_k(\xi)$ among the nodes. Fix m, x, our estimations (3.26) and (3.30) suggest that error $|\Delta_k(\xi)| = |\hat{\hat{a}}(\frac{\xi+k}{m}) - \hat{a}(\frac{\xi+k}{m})|$ in the worst possible case could be proportional to $\delta_k(\xi)\delta^2(\xi)$. Our numerical experiment suggests this is sharp, see Figure 2 (c) 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 by the proof of Theorem 3.4.5, $||\mathbf{H}_m^{-1}||_\infty$ grows at least geometrically when m increases. Thus, when m increases, the infinity norm of $\mathbf{H}_m^{-1}(\xi)$ gets bigger and our solutions become more likely less robust to noise, see Figure 2 (a) and (b).

3.5 Numerical Experiment

In this section, we provide some simple numerical simulations to verify some theoretical accuracy estimations in section 3.4.

3.5.1 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 $\hat{a}(\xi) = 0.1 + 0.8\cos(2\pi\xi) + 0.1\cos(4\pi\xi)$, $x = (\cdots, 0, 0.242, 0.383, 0.242, 0, \cdots)$ such that $\hat{x}(\xi) = 0.383 + 0.484\cos(2\pi\xi)$. We 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 $\widehat{\tilde{y}_l}(\xi_i) = \hat{y}_l(\xi_i) + \varepsilon_l$ for $l = 0, \cdots, 5$, where y_l is defined as in (3.1) and $\varepsilon_l \sim 10^{-10}$.

2. Use Algorithm 3.3.1 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. Choose $\xi = 0.3$ and $m = 2 : 1 : 7$, we compute $||\mathbf{H}_m^{-1}(0.3)||_\infty$ for different m.

3.5.2 Experiment Results

In this subsection, we plot several figures to reflect the experiment results. The x-axis of the Figure 2 (a)$-$(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 $\xi_1, \cdots, \xi_9$ are more and more closer to $\frac{1}{2}$, we expect the infinity norm of $\mathbf{H}_m^{-1}(\xi)$ to get sufficiently larger and larger. Note that 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 Figure 2 (a) and (b). They exhibit almost the same behaviour and grows proportionally. 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 estimation** (3.19) **and** (3.28). Our estimation (3.19) and (3.28) suggest that 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 Figure 2 (c) 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 estimation(3.19) and (3.28). It is worthy 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 the quantity is, the less accuracy the Algorithm is.

3. **The infinity norm of $\mathbf{H}_m^{-1}(\xi)$.** Recall in this experiment, we choose $m = 2,3,\cdots,6,7$ and $\xi = 0.3$. We plot the value of $||\mathbf{H}_m^{-1}(0.3)||_\infty$ for different m. The results are presented in Figure 2 (f). The $y-$axis is set to be logarithmic. It is shown that $||\mathbf{H}_m^{-1}(\xi)||_\infty$ grows geometrically.

3.6 Other Numerical Mehtods

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 to our case. These two classical methods are well known for their better numerical stability than the original Prony method.

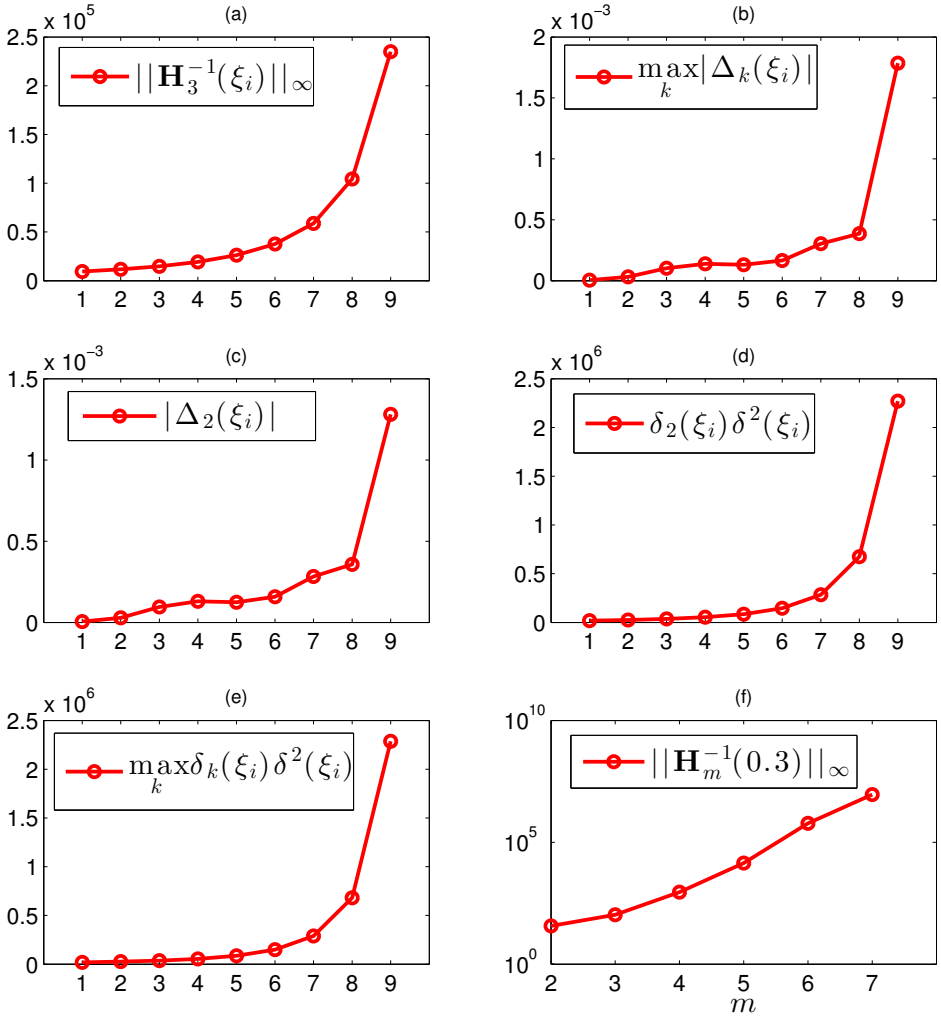

Figure 3.2: Experiment Results

3.6.1 Generalized Matrix Pencil Method

Let L and N be two integers satisfying $L \geq m$ and $N \geq L+m$. Similarly, 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

$$\mathbf{H}^{\xi}_{N-L,L+1} = \big[\mathbf{h}_0(\xi), \mathbf{h}_1(\xi), \cdots, \mathbf{h}_L(\xi)\big], \tag{3.33}$$

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

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

$$\mathbf{H}^{\xi}_{N-L,L}(s) = V_{m,N-L}(\mathbf{w}(\xi))^T diag(\mathbf{x}(\xi)) diag(\mathbf{w}(\xi))^s V_{m,L}(\mathbf{w}(\xi)). \tag{3.34}$$

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

Lemma 3.6.1. *Let N, L be two postive integers s.t. $m \leq L \leq N-m$. Assume $\xi \neq 0, \frac{1}{2}$ and $diag(\mathbf{x}(\xi))$ is invertible. The solutions to the generalized singular eigenvalue problem :*

$$(z\mathbf{H}^{\xi}_{N-L,L}(0) - \mathbf{H}^{\xi}_{N-L,L}(1))\mathbf{p}(\xi) = 0 \tag{3.35}$$

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 can be done by the factorization (3.34) and a similar manner with the proof of Theorem 2 in [31]. □

Proposition 3.6.2. *Let* N, L *be two postive integers s.t.* $m \leq L \leq N-m$. *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 (3.35) by $\mathbf{H}^{+}{}^{\xi}_{N-L,L}(0)$, we have

$$\mathbf{H}^{+}{}^{\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)\mathbf{p}_i(\xi) = z_i\mathbf{H}^{+}{}^{\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(0)\mathbf{p}_i(\xi), \tag{3.36}$$

By property of pseudoinverse, $\mathbf{H}^{+}{}^{\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(0)$ is the orthogonal projection onto the $\mathbf{R}(\mathbf{H}^{*}{}^{\xi}_{N-L,L}(0))$. 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 $\mathbf{H}^{+}{}^{\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$ is of rank $m \leq L$, $\mathbf{H}^{+}{}^{\xi}_{N-L,L}(0)\mathbf{H}^{\xi}_{N-L,L}(1)$ has $L-m$ zero eigenvalues. □

It is immediate to see 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 just need to 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 idea of Algorithm 5 for SVD based Matrix Pencil Method in [32], we can employ the Singular Value Decomposition(SVD) of the Hankel matrices.

Lemma 3.6.3. *In addition to the conditions of Proposition 3.6.2, 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 computations 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 3.6.1. Note that the amount of computation required by Algorithm 3.6.1 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 good choice for L is between one third of N and two thirds of N [32]. In our numerical example, we choose L to be around one third of N.

Algorithm 3.6.1 Generalized Matrix Pencil Method (Based on SVD)

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

1: Compute the Fourier transformation of the measurement sequences $\{y_l\}_{l=0}^{N-1}$ and 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 rest eigenvalues by the monotonicity and symmetric condition of $\hat{a}$ to get $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

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

3.6.2 Generalized ESPRIT Method

Original ESPRIT Method relies on a particular property of Vandermonde matrices known as the rotational invariance [41]. By the factorization results (3.34), 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 Vandermonde matrix generated by $\{\hat{a}(\frac{\xi+i}{m}), i = 0, \cdots, m-1\}$. Hence we can generalize the idea and present the generalized ESPRIT algorithm based on SVD for estimating the $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$ in our case. We summarize it in Algorithm 3.6.2

Algorithm 3.6.2 Generalized ESPRIT Algorithm

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

1: Compute the Fourier transformation 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 symmetric condition of $\hat{a}$ to get $\{\hat{a}(\frac{\xi+i}{m}) : i = 0, \cdots, m-1\}$.

5:

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

3.6.3 Data Preprocessing Using Cadzow Denoising Method

It has been shown in the previous sections 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 hypothesis:

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 [11] to preprocess the noisy data and guarantee to build a Hankel matrix with above two key properties. In our context, it can be summarized in Algorithm 3.6.3.

The procedure of Algorithm 3.6.3 is guaranteed to converge to a matrix which exhibits the desired two key properties [11]. 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 3.3.1 does not perform well when noise is big, we can combine the Algorithm 3.3.1 and Algorithm 3.6.3 to recover the Fourier spectrum of a and improve the performance. In our numerical example, we choose $L = m$.

Algorithm 3.6.3 Cadzow Iterative Denoising Method

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

1: Build the Hankel matrix $\mathbf{H}^{\xi}_{N-L,L+1}$ from $\{\hat{y}_l(\xi)\}_{l=0}^{N-1}$ and preform 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** $\frac{\lambda_{m+1}}{\lambda_m} \geq \varepsilon$ **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 $\{\hat{y}_l(\xi)\}_{l=0}^{N-1}$ from the first column and the last row of $\mathbf{H}^{\xi}_{N-L,L+1}$

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

3.7 Numerical Examples

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

Example 3.7.1. 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 $\varepsilon_l \sim U(-\varepsilon, \varepsilon)$ to the Fourier data $\hat{y}_l$ for $l = 0, \cdots, N-1$. Recall 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 $k = 0, 1, m-1$. 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 $\mathbf{e}_{worst}(\xi) = \max\limits_k \mathbf{e}_k(\xi)$. Besides, we define the mean square error

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

Then we apply our Algorithm 3.3.1, Algorithms 3.3.1+Algorithm 3.6.3, Algorithm 3.6.1 and Algorithm 3.6.2 to the case when $\varepsilon = 0.4$. For several parameters N and L, the resulting errors (average over 100 experiments) are presented in Table 3.1. As the bound ε in the algorithms we use 10^{-10}. It is shown in the table that increasing the temporal samples, i.e. N, will help reduce the error. The new proposed algorithms have better performance than Algorithm 3.3.1, if given more spatiotemporal data.

3.8 Concluding Remarks

In this chapter, we have investigated 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 (3.1), where $Ax = a * x$. We show that if one doubles the amount of temporal samples needed in [6] 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 the finite impulse response filter and signal, if an upper bound for their support is known. In particular, we have done a first order perturbation analysis and the estimates are formulated in very simple geometric terms involving Fourier spectral function of a, x and m, shedding some light on the structure of the problem. We get a lower bound estimation for infinity

Table 3.1: Numerical Results

Algorithm	m	N	L	$\mathbf{e}_{best}$	$\mathbf{e}_{worst}$	**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
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
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
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

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 the robustness to noise. The potential applications includes the One-Chip Sensing: sensors inside chips for accurate measurements of voltages, currents, and temperature (e.g., avoid overheating any area of the chip), sources localization of an evolving state and time-space trade off(e.g., sound field acquisition using microphones)etc.

Chapter 4

MULTIDIMENSIONAL SPATIOTEMPORAL TRADE OFF PROBLEM IN DISCRETE INVARIANT EVOLUTION SYSTEMS

4.1 Problem Formulation

In this section, we formulate the spatiotemporal trade off problem in the multidimensional invariant evolution system. Let f be an initial state definied on the multidimensional lattice $D = \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}$ and $\mathbb{Z} \times \mathbb{Z}$. At time instance $t = n \in \mathbb{N}$, the initial state f is altered by convolution with a filter $a \in \ell^1(D)$ n times to be $A_n(f) = a * a * ... * a * f = a^n * f$. At each time instance $t = n$, the altered state $A_n(f)$ is under-sampled at a uniform subsampling rate m_1 and m_2 in each direction. Namely we take spatiotemporal samples on a uniform sublattice X of D. Let S_X be the subsampling operator defined by $S_X(f) = f(X)$. We ask the following question:

Problem 4.1.1. Under what conditions on a, X and N that can we recover any initial state $f \in \ell^2(D)$ from the spatiotemporal samples

$$\{f(X),\, a * f(X),\, \cdots,\, (a^{N-1} * f)(X)\},\ \text{for}\ X \subseteq D? \tag{4.1}$$

The above problem is solved when conditions on the sampling sets and time instances N are found, such that recovery of the signal is possible, preferably in a stable way.

4.2 Previous Work

In [5, 6] the authors have studied the spatiotemporal trade off problem for the discrete spatially invariant evolution system, in which the initial state f is defined on the domain $D = \mathbb{Z}_d$ and $\mathbb{Z}$. At each time instance $t = n$, the altered state $A_n(f)$ is under-sampled at

a uniform subsampling rate m. The invertibility and stability questions have been fully answered under the specific constraints of the convolution operator A. The multidimensional spatiotemporal trade off problem we consider in this chapter also has similarities with problems considered by some other authors. For example, in [1], the authors work in a multivariable shift-invariant space(MSIS) setting, and study linear systems $\{L_j : j = 1,\cdots,s\}$ such that one can recover any f in MSIS by uniformly downsampling the functions $\{(L_j f) : j = 1,\cdots,s\}$, i.e. taking the generalized samples $\{(L_j f)(M\alpha)\}_{\alpha\in\mathbb{Z}_d, j=1,\cdots,s}$. In dynamical sampling, there is only one convolution operator A, and it is applied iteratively to the function f. This iterative structure is important for our analysis of the kernel of the arising matrix, and using that special structure we are able to add extra samples outside of the initial uniform sampling grid and get full recovery of the signal.

4.3 Contribution and Organization

Our goal is to extend the one variable results in [5, 6] to the multidimensional setting. In section 4.4, we consider the finite dimensional case $D = \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}$. We derive the conditions on the convolution kernel a such that 1 to 1 spatiotemporal trade off rate can be achieved. However, for the cases when the convolution operator has symmetries in the Fourier domain, uniform sampling X is not enough to achieve the stable recovery for all initial signal. We successfully overcome this singularity problem by adding some extra spatial samples of f. As we will see later, the two variable problem is more complicated in structure and we find it more subtle to overcome the singularity problems. In section 4.5, we study the infinite dimensional case. Studying the stated Problem 4.1.1 in higher variable setting would require similar techniques to the ones we use in this chapter to expand the domain from one to two dimensions.

4.4 Case I: $D = \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}$

For a positive integer d, $\mathbb{Z}_d$ denotes the finite group of integers modulo d. In the finite discrete setting, we work on the domain $D = \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}$, $d_1, d_2 \in \mathbb{N}^+$. Let the operator A act on the signal of interest $f \in \ell^2(D)$ as a convolution with some $a \in \ell^1(D)$ given by

$$Af(k,l) = a * f(k,l) = \sum_{(s,p)\in D} a(s,p) f(k-s, l-p), \text{ for all } (k,l) \in D. \tag{4.2}$$

Note that A is a bounded linear operator that maps $\ell^2(D)$ to itself. The initial signal f is evolving in time under the repeated effect of A such that at time instance $t = n$, the evolved signal is $f_n = A^n f = a * a * \cdots * a * f$ (and $f = f_0 = A^0 f$).

We assume that d_1 and d_2 are odd numbers, such that $d_i = J_i m_i$ for integers $m_i \geq 1$, $J_i \geq 1$, $i = 1, 2$. We set the sampling sensors on a uniform coarse grid $X = m_1\mathbb{Z}_{d_1} \times m_2\mathbb{Z}_{d_2}$ to sample the initial state f and its temporally evolved states Af, $A^2 f, \ldots, A^{N-1} f$. Note that, given such a coarse sampling grid, each individual measurement is insufficient for recovery of the sampled state.

Let $S_X = S_{m_1,m_2}$ denote the assigned subsampling operator related to the sampling grid. Specifically,

$$(S_X f)(k,l) = \begin{cases} f(k,l) & \text{if } (k,l) \in X \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

For some $N \geq 2$, our objective is to reconstruct f from the combined coarse samples set

$$\{y_j = S_X(A^j f)\},\ j = 0, 1, ..., N-1. \tag{4.4}$$

We denote by $\mathscr{F}$ the $2-$dimensional discrete Fourier transform (2d DFT) and use the notation $\hat{x} = \mathscr{F}(x)$. After applying $\mathscr{F}$ to (4.4), due to the two-dimensional Poisson's summation

formula, we obtain

$$\hat{y}_n(i,j) = \frac{1}{m_1 m_2} \sum_{k=0}^{m_1-1} \sum_{l=0}^{m_2-1} \hat{a}^n(i+kJ_1, j+lJ_2)\, \hat{f}(i+kJ_1, j+lJ_2) \tag{4.5}$$

for $(i,j) \in \mathscr{I} = \{0,\cdots,J_1-1\} \times \{0,\cdots,J_2-1\}$ and $n = 0,1,\ldots,N-1$.

Let $\bar{\mathbf{y}}(i,j) = (\hat{y}_0(i,j)\ \ \hat{y}_1(i,j)\ \ \ldots\ \ \hat{y}_{N-1}(i,j))^T$, $(i,j) \in \mathscr{I}$ and

$$\bar{\mathbf{f}}(i,j) = \begin{pmatrix} \hat{f}(i,j) \\ \hat{f}(i+J_1,j) \\ \vdots \\ \hat{f}(i+(m_1-1)J_1,j) \\ \hat{f}(i,j+J_2) \\ \vdots \\ \hat{f}(i+(m_1-1)J_1,j+J_2) \\ \vdots \\ \vdots \\ \hat{f}(i,j+(m_2-1)J_2) \\ \vdots \\ \hat{f}(i+(m_1-1)J_1,j+(m_2-1)J_2) \end{pmatrix}.$$

We use the block-matrices

$$A_{l,m_1m_2}(i,j) = \begin{pmatrix} \overset{1}{\hat{a}(i,j+lJ_2)} & \overset{1}{\hat{a}(i+J_1,j+lJ_2)} & \overset{\ldots}{\ldots} & \overset{1}{\hat{a}(i+(m_1-1)J_1,j+lJ_2)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{a}^{N-1}(i,j+lJ_2) & \hat{a}^{N-1}(i+J_1,j+lJ_2) & \ldots & \hat{a}^{N-1}(i+(m_1-1)J_1,j+lJ_2) \end{pmatrix},$$

where $l = 0,1,...,m_2-1$, to define the $N \times m_1m_2$ matrix

$$\mathscr{A}_{m_1,m_2}(i,j) = [A_{0,m_1m_2}(i,j)\, A_{1,m_1m_2}(i,j) \ldots A_{m_2-1,m_1m_2}(i,j)] \tag{4.6}$$

for all $(i,j) \in \mathscr{I}$. Equations (4.5) have the form of vector inner products, so we restate

them in matrix product form

$$\bar{\mathbf{y}}(i,j) = \frac{1}{m_1 m_2} \mathscr{A}_{m_1,m_2}(i,j)\bar{\mathbf{f}}(i,j). \tag{4.7}$$

By equation (4.7), we need $N \geq m_1 m_2$ to be able to recover the signal f. Note that for $N = m_1 m_2$, matrix (4.6) is square, we denote this special square matrix by $A_{m_1,m_2}(i,j)$ and obtain the following reconstruction result:

Proposition 4.4.1. *For $N = m_1 m_2$, we can recovery any f from spatiotemporal samples defined in* (4.4) *in a stable way if and only if*

$$\det A_{m_1,m_2}(i,j) \neq 0 \, \textit{for all } (i,j) \in \mathscr{I}. \tag{4.8}$$

Note that in the finite dimensional case, unique reconstruction is equivalent to stable reconstruction. When (4.8) holds true, the signal is recovered from the system of equations

$$\bar{\mathbf{f}}(i,j) = m_1 m_2 A^{-1}_{m_1,m_2}(i,j)\bar{\mathbf{y}}(i,j), \ (i,j) \in \mathscr{I}.$$

As expected, Proposition 4.4.1 reduces to the respective result in [5] when $d = d_1$ and $d_2 = 1$, or $d = d_2$ and $d_1 = 1$.

4.4.1 Extra samples for stable spatiotemporal sampling

Proposition 4.4.1 gives a complete characterization of stable recovery from the dynamical samples (4.4). In practice, however, we may not have the ideal filter a such that (4.8) holds true. For instance, consider a kernel a with a so-called *quadrantal symmetry*, i.e. let

$$\hat{a}(s,p) = \hat{a}(d_1 - s, p) = \hat{a}(s, d_2 - p) = \hat{a}(d_1 - s, d_2 - p)$$

for all $(s,p) \in D$. Since (4.6) is a Vandermonde matrix, it is singular if and only if some of its columns coincide. In this case, it is easy to see that $A_{m_1,m_2}(0,0)$ is singular, which prevents us from the stable reconstruction. In particular, for this special case, no matter how many times we take temporal samples on the uniform grid, we don't gain any new information. This fact is due to the special structure of Vandermonde matrix. Hence some spatiotemporal samples of the evolving signal must be taken on extra spatial locations to overcome the lack of stability. The problem is, in what way we can take extra spatiotemporal samples? Can we give a complete characterization to these successful candidates?

Motivated by approach in [5], here we also propose a way of taking extra spatial samples of f at initial time level to overcome the lack of reconstruction uniqueness, whenever singularities for matrix (4.6) occur. Note that once the uniqueness is achieved, then stability of reconstruction is also achieved, by the finite dimensional nature of this problem. Let us assume

$$\mathscr{A} = \begin{pmatrix} A_{m_1,m_2}(0,0) & 0 & \dots & 0 \\ 0 & A_{m_1,m_2}(1,0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{m_1,m_2}(J_1-1,J_2-1) \end{pmatrix}$$

and

$$\bar{\mathbf{f}} = \begin{pmatrix} \bar{\mathbf{f}}(0,0) \\ \bar{\mathbf{f}}(1,0) \\ \vdots \\ \bar{\mathbf{f}}(J_1-1,0) \\ \bar{\mathbf{f}}(1,1) \\ \vdots \\ \bar{\mathbf{f}}(J_1-1,1) \\ \vdots \\ \vdots \\ \bar{\mathbf{f}}(0,J_2-1) \\ \vdots \\ \bar{\mathbf{f}}(J_1-1,J_2-1) \end{pmatrix}, \bar{\mathbf{y}} = \begin{pmatrix} \bar{\mathbf{y}}(0,0) \\ \bar{\mathbf{y}}(1,0) \\ \vdots \\ \bar{\mathbf{y}}(J_1-1,0) \\ \bar{\mathbf{y}}(1,1) \\ \vdots \\ \bar{\mathbf{y}}(J_1-1,1) \\ \vdots \\ \vdots \\ \bar{\mathbf{y}}(0,J_2-1) \\ \vdots \\ \bar{\mathbf{y}}(J_1-1,J_2-1) \end{pmatrix}.$$

Then

$$\mathscr{A}\bar{\mathbf{f}} = \bar{\mathbf{y}} \tag{4.9}$$

and

$$\ker(\mathscr{A}) = \bigoplus_{(i,j)\in\mathscr{I}} \ker[A_{m_1,m_2}(i,j)]. \tag{4.10}$$

The kernels of each $A_{m_1,m_2}(i,j)$ can be viewed as generated by linearly independent vectors $\hat{v}_j \in \ell^2(D)$ such that each $\hat{v}_j$ has exactly two nonzero coordinates, one of which is equal to 1 and the other is -1. Let's assume that the nullity of matrix $A_{m_1,m_2}(i,j)$ equals $w_{i,j}$ at each $(i,j) \in \mathscr{I}$. Then there are $n = \sum_{i,j} w_{i,j}$ of such linearly independent vectors $\hat{v}_j \in \ell^2(D)$. Let $\{v_j : j = 1,\cdots,n\}$ be their image under the $2D$ inverse DFT. Note that $\{v_j : j = 1,\cdots,n\} \subseteq \ell^2(D)$ is also linearly independent.

Let $\Omega \subseteq D \setminus X$ be the additional sampling set, that is to say, we take extra spatial samples of the initial state f at the locations specified by Ω. By S_Ω we denote the related sampling operator and R_Ω is a $|\Omega| \times n$ matrix with rows corresponding to $[v_1(k,l),\cdots,v_n(k,l)]_{\{(k,l)\in\Omega\}}$.

With these notations, the following result holds true:

Theorem 4.4.2. *The reconstruction of $f \in \ell^2(D)$ from its spatiotemporal samples*

$$\{S_\Omega f, S_X f, S_X A f, \cdots, S_X A^{m_1 m_2 - 1} f\} \tag{4.11}$$

is possible in a stable manner if and only if $rank(R_\Omega) = n$.

In particular, if SSP holds true, then we must have $|\Omega| \geq n$.

Proof. Let $W = span\{v_j : j = 1, \cdots, n\}$. It suffices to show that

$$ker(S_\Omega) \cap W = \{0\} \text{ if and only if } rank(R_\Omega) = n.$$

Suppose w is in $ker(S_\Omega) \cap W$. There must exist coefficients $c_1, c_2, .., c_n$ so that $w = \sum_{j=1}^{n} c_j v_j$ and $S_\Omega w = 0$. The last statement is equivalent to

$$[v_1(k,l), v_2(k,l), \cdots, v_n(k,l)] \, [c_1 \; c_2 \; ... \; c_n]^T = 0$$

for each $(k,l) \in \Omega$. Equivalently, we have $R_\Omega \mathbf{c} = 0$. Hence, $\mathbf{c} = 0$ if and only if $rank(R_\Omega) = n$. □

Since the $d_1 d_2 \times n$ matrix $R = [v_1(k,l), \cdots, v_n(k.l)]_{\{(k,l) \in D\}}$ has column rank n, for any kernel a, there exists a minimal choice of Ω, namely $|\Omega| = n$ such that the square matrix R_Ω is invertible. It is hard to give a formula to specify the extra sampling set for every kernel $a \in \ell^2(D)$. On the other hand, compared to the $1-$variable case [5], it is more challenging to specify the rank of R_Ω analytically, since the entries of R_Ω will involve the product of sinusoids mixed with exponentials in general.

In [5], the authors studied a typical low pass filter with symmetric properties and gave a choice of a minimal extra sampling set Ω, since symmetry reflects the fact that there is often no preferential direction for physical kernels and monotonicity is a reflection of

energy dissipation. Similarly, we consider a kernel a with a so-called *strict quadrantal symmetry*: for a fixed $(k,l) \in D$, $\hat{a}(s,p) = \hat{a}(k,l)$ if and only if

$$(s,p) \in \{(k,l),(d_1-k,l),(k,d_2-l),(d_1-k,d_2-l)\}. \tag{4.12}$$

Since $A_m(i,j)$ is a Vandermonde matrix, it has singularity if and only if some of its columns coincide. We can compute the singularity of each $A_m(i,j)$, as we make use of its special structure.

Lemma 4.4.3. *If the filter a satisfies the symmetry assumptions* (4.12)*, then*

$$dim(ker(\mathscr{A})) = \frac{d_1(m_2-1)}{2} + \frac{d_2(m_1-1)}{2} - \frac{(m_1-1)(m_2-1)}{4}.$$

Clearly, we need an extra sampling set $\Omega \subseteq D$ with size $dim(ker(\mathscr{A}))$. Based on Theorem 4.4.2, we provide a minimal Ω:

Theorem 4.4.4. *Assume that the kernel a satisfies the strict quadrantal symmetry assumptions* (4.12) *and let*

$$\Omega = \{(k,l) : k = 1 \cdots \frac{m_1-1}{2}, l \in \mathbb{Z}_{d_2}\} \cup \{(k,l) : k \in \mathbb{Z}_{d_1}, l = 1, \cdots, \frac{m_2-1}{2}\}.$$

Then, any $f \in \ell^2(D)$ is recovered in a stable way from the expanded set of samples

$$\{S_\Omega f, S_X f, S_X A f, \cdots, S_X A^{m_1 m_2 - 1} f\}. \tag{4.13}$$

Remark 4.4.5. Note that in this case

$$|\Omega| = \frac{d_1(m_2-1)}{2} + \frac{d_2(m_1-1)}{2} - \frac{(m_1-1)(m_2-1)}{4},$$

so by Theorem 4.4.2 and Lemma 4.4.3 we can not do better in terms of its cardinality.

Proof. Set

$$n = \frac{d_1(m_2-1)}{2} + \frac{d_2(m_1-1)}{2} - \frac{(m_1-1)(m_2-1)}{4}.$$

Recall that the kernels of singular blocks $A_{m_1,m_2}(i,j)$ are generated by vectors $\{\hat{v}_k : k = 1,\cdots n\}$, such that each $\hat{v}_k$ has exactly two non-zero components, 1 and -1 (corresponding to each pair of identical columns). Then the formula of $2D$ inverse DFT gives

$$v_j(k,l) = \sum_{s=0}^{d_1-1}\sum_{p=0}^{d_2-1} \hat{v}_j(s,p) e^{\frac{2\pi isk}{d_1}} e^{\frac{2\pi ipl}{d_2}}, \ \ (k,l) \in \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}. \tag{4.14}$$

We define a row vector $\mathscr{F}_1(k) = \left[1, e^{\frac{2\pi ik}{d_1}}, \cdots, e^{\frac{2\pi i(d_1-1)k}{d_1}}\right]$ for all $k \in \mathbb{Z}_{d_1}$. For each $l = 0,1,\cdots,d_2-1$, we define a row vector $\bar{\mathscr{F}}_2(l)$ of length $d_2 - \frac{m_2-1}{2}$, which is derived from vector

$$[1, e^{\frac{2\pi il}{d_2}}, \cdots, e^{\frac{2\pi i(d_2-1)l}{d_2}}]$$

after deleting the entries that correspond to $\{sJ_2+1 : \ 1 \le s \le \frac{m_2-1}{2}\}$, i.e. we omit the entries $e^{\frac{2\pi sJ_2}{d_2}}$ for $1 \le s \le \frac{m_2-1}{2}$. We reorder the vectors v_j so that $[v_1(k,l),\cdots,v_n(k,l)]$ equals

$$2i\left[\sin(\tfrac{2\pi 1l}{m_2})\mathscr{F}_1(k),\ \cdots,\ \sin(\tfrac{2\pi(m_2-1)l}{2m_2})\mathscr{F}_1(k),\ \sin(\tfrac{2\pi 1k}{m_1})\bar{\mathscr{F}}_2(l)\ \cdots,\ \sin(\tfrac{2\pi(m_1-1)k}{2m_1})\bar{\mathscr{F}}_2(l)\right]$$

for every $(k,l) \in \Omega$. By Theorem 4.4.2, the proof is complete if we show that these $n = |\Omega|$ row vectors of size n are linearly independent.

We define a row vector $R(k,l)$ corresponding to $(k,l) \in \Omega$ given by

$$2i\left[\sin(\tfrac{2\pi 1l}{m_2})\mathscr{F}_1(k),\ \cdots,\ \sin(\tfrac{2\pi(m_2-1)l}{2m_2})\mathscr{F}_1(k),\ \sin(\tfrac{2\pi 1k}{m_1})\bar{\mathscr{F}}_2(l)\ \cdots,\ \sin(\tfrac{2\pi(m_1-1)k}{2m_1})\bar{\mathscr{F}}_2(l)\right].$$

Suppose that for some coefficients $\{c(k,l) : (k,l) \in \Omega\}$, it holds

$$\sum_{(k,l)\in\Omega} c(k,l)R(k,l) = 0.$$

We need to show that all $c(k,l) = 0$. Note that, for a fixed k, the vector $R(k,l)$ is compartmentalized into two components with lengths $\frac{m_2-1}{2}$ and $\frac{m_1-1}{2}$. By construction, $\{\mathscr{F}_1(k) \mid k \in \mathbb{Z}_{d_1}\}$ are linearly independent row vectors. Then, the coefficients related to $\mathscr{F}_1(k)$ for the first component should be zeros. Related to the first component of length $\frac{m_2-1}{2}$, for every fixed $k \in \mathbb{Z}_{d_1}$ such that $(k,l) \in \Omega$ for some l, the following m_2 equations hold true

$$\sum_{(k,l)\in\Omega} c(k,l)\sin\left(\frac{2\pi sl}{m_2}\right) = 0 \text{ for } s = 0,1,...,m_2-1. \tag{4.15}$$

Case I if $k \geq \frac{m_1+1}{2}$ or $k = 0$, then $(k,l) \in \Omega$ if and only if $l = 1,\cdots\frac{m_2-1}{2}$. We restate the system of equations (4.15) in the matrix form:

$$\begin{pmatrix} \sin(\frac{2\pi}{m_2}) & \sin(\frac{4\pi}{m_2}) & \dots & \sin(\frac{\pi(m_2-1)}{m_2}) \\ \sin(\frac{4\pi}{m_2}) & \sin(\frac{8\pi}{m_2}) & \dots & \sin(\frac{2\pi(m_2-1)}{m_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\frac{\pi(m_2-1)}{m_2}) & \sin(\frac{2\pi(m_2-1)}{m_2}) & \dots & \sin(\frac{\pi(m_2-1)(m_2-1)}{2m_2}) \end{pmatrix} \begin{pmatrix} c(k,1) \\ c(k,2) \\ \vdots \\ c(k,\frac{m_2-1}{2}) \end{pmatrix} = \mathbf{0}.$$

The matrix on the left-hand side is invertible, since

$$\{sin(2\pi x), sin(4\pi x), ..., sin((m_2-1)\pi x)\}$$

is a Chebyshev system on $[0,1]$(see[**?**]); Hence we have $c(k,l) = 0$ for $l = 1,\cdots\frac{m_2-1}{2}$.

Case II if $1 \leq k \leq \frac{m_1-1}{2}$, then $(k,l) \in \Omega$ if and only if $l = 0,\cdots,d_2-1$. Then (4.15) is

equivalent to the system of equations

$$\sum_{l=0}^{d_2-1} c(k,l)\sin\left(\frac{2\pi sl}{m_2}\right) = 0 \text{ for } s = 1,2,...,(m_2-1)/2. \tag{4.16}$$

Related to the second component of length $\frac{m_1-1}{2}$, and combined with the fact that $c(k,l) = 0$ if k is in case I, for all $s = 1,2,...,\frac{m_1-1}{2}$ we have

$$\sum_{l=0}^{d_2-1}\left(\sum_{k=1}^{\frac{m_1-1}{2}} c(k,l)\sin\left(\frac{2\pi sk}{m_1}\right)\bar{\mathscr{F}}_2(l)\right) = 0. \tag{4.17}$$

Let $\bar{\mathscr{F}}_2 = [\bar{\mathscr{F}}_2(0)^T,\cdots,\bar{\mathscr{F}}_2(d_2-1)^T]$, where $\bar{\mathscr{F}}_2(l)^T$ denotes the transpose of each row vector $\bar{\mathscr{F}}_2(l)$; $\bar{\mathscr{F}}_2$ is a $(d_2 - \frac{m_2-1}{2}) \times d_2$ matrix. Using matrix notation, the first equation in (4.17) can be restated as a product, namely

$$\bar{\mathscr{F}}_2 \cdot \begin{pmatrix} \sum_{k=1}^{\frac{m_1-1}{2}} \sin\left(\frac{2\pi k}{m_1}\right) c(k,0) \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin\left(\frac{2\pi k}{m_1}\right) c(k,1) \\ \vdots \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin\left(\frac{2\pi k}{m_1}\right) c(k,d_2-1) \end{pmatrix} = 0$$

As an easy consequence of equation (4.16), for each $1 \le j \le \frac{m_2-1}{2}$, it holds

$$\sum_{k=1}^{\frac{m_1-1}{2}} \sin\left(\frac{2\pi k}{m_1}\right)\sum_{l=0}^{d_2-1}\left(\sin(\frac{2\pi lj}{m_2})c(k,l)\right) = 0, \tag{4.18}$$

which is equivalent to

$$\sum_{k=1}^{\frac{m_1-1}{2}}\sum_{l=0}^{d_2-1} \sin\left(\frac{2\pi lj}{m_2}\right)\sin\left(\frac{2\pi k}{m_1}\right) c(k,l) = 0,$$

$$\text{i.e. } \sum_{l=0}^{d_2-1} \sin\left(\frac{2\pi l j}{m_2}\right) \sum_{k=1}^{\frac{m_1-1}{2}} \sin\left(\frac{2\pi k}{m_1}\right) c(k,l) = 0. \tag{4.19}$$

We define a $\frac{m_2-1}{2} \times d_2$ matrix E as follows:

$$E = \begin{pmatrix} \sin(\frac{2\pi\cdot 0}{m_2}) & \sin(\frac{2\pi\cdot 1}{m_2}) & \dots & \sin(\frac{2\pi(d_2-1)}{m_2}) \\ \sin(\frac{4\pi\cdot 0}{m_2}) & \sin(\frac{4\pi * \cdot 1}{m_2}) & \dots & \sin(\frac{4\pi(d_2-1)}{m_2}) \\ \vdots & \vdots & \dots & \vdots \\ \sin(\frac{\pi(m_2-1)\cdot 0}{m_2}) & \sin(\frac{2\pi(m_2-1)}{m_2}) & \dots & \sin(\frac{\pi(m_2-1)(d_2-1)}{m_2}) \end{pmatrix}.$$

Due to (4.19), we have

$$E \cdot \begin{pmatrix} \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,0) \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,1) \\ \vdots \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,d_2-1) \end{pmatrix} = \mathbf{0}. \tag{4.20}$$

Let $\mathscr{F}_2 = \begin{pmatrix} E \\ \bar{\mathscr{F}}_2 \end{pmatrix}$. Then

$$\mathscr{F}_2 \cdot \begin{pmatrix} \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,0) \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,1) \\ \vdots \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,d_2-1) \end{pmatrix} = \mathbf{0}.$$

Note that the $d_2 \times d_2$ matrix $\mathscr{F}_2$ is invertible, since it is the image of a series of

elementary matrices acting on the $d_2 \times d_2$ DFT matrix (one row minus another row).

Hence we have

$$\begin{pmatrix} \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,0) \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi k}{m_1}) c(k,1) \\ \vdots \\ \sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{\pi(m_1-1)k}{m_1}) c(k,d_2-1) \end{pmatrix} = \mathbf{0}. \tag{4.21}$$

After analyzing the rest of the equations in (4.17), we obtain:

$$\sum_{k=1}^{\frac{m_1-1}{2}} \sin(\frac{2\pi jk}{m_1}) c(k,s) = 0 \text{ for } j = 2, \cdots, \frac{m_1-1}{2}, \ s = 0,1,...,d_2-1.$$

In a similar manner, for each $l = 0, \cdots, d_2 - 1$ we obtain the matrix equation

$$\begin{pmatrix} \sin(\frac{2\pi}{m_1}) & \sin(\frac{4\pi}{m_2}) & \dots & \sin(\frac{\pi(m_1-1)}{m_1}) \\ \sin(\frac{4\pi}{m_1}) & \sin(\frac{8\pi}{m_2}) & \dots & \sin(\frac{2\pi(m_1-1)}{m_1}) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\frac{\pi(m_1-1)}{m_1}) & \sin(\frac{2\pi(m_1-1)}{m_1}) & \dots & \sin(\frac{\pi(m_1-1)(m_1-1)}{2m_1}) \end{pmatrix} \begin{pmatrix} c(1,l) \\ c(2,l) \\ \vdots \\ c(\frac{m_1-1}{2},l) \end{pmatrix} = 0.$$

As the matrix on the left hand side is invertible, we must have $c(k,l) = 0$ for $k = 1, \cdots, \frac{m_1-1}{2}$.

We have demonstrated that $c(k,l) = 0$ for all $(k,l) \in \Omega$. Therefore the n row vectors $\{R(k,l)\}_{(k,l)\in\Omega}$ are linearly independent i.e. stability of the signal recovery is achieved. $\square$

4.5 Case II: $\mathbb{Z} \times \mathbb{Z}$

In this section, we aim to generalize our results to signals of infinite length. Somewhat surprisingly, there is not much difference between the techniques used in these two settings and we feel that we can gloss over a few details in the second part without overburdening the reader.

Let $D = \mathbb{Z} \times \mathbb{Z}$. We study a signal of interest $f \in \ell^2(D)$ that evolves over time under the influence of an evolution operator A. The operator A is described by a convolution with $a \in \ell^1(D)$, namely

$$Af(p,q) = a * f(p,q) = \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} a(k,l) f(p-k,q-l) \text{ at all } (p,q) \in D.$$

Clearly, A is a bounded linear operator, mapping $\ell^2(D)$ to itself. Given integers m_1,$m_2 \geq 1$, we assume m_1 and m_2 are odd number. We introduce a coarse sampling grid $X = m_1\mathbb{Z} \times m_2\mathbb{Z}$. We make use of a uniform sampling operator S_X, defined by $(S_X f)(k,l) = f(m_1 k, m_2 l)$ for $(k,l) \in D$. The goal is to reconstruct f from the set of coarse samples

$$\begin{cases} y_0 = S_X f \\ y_1 = S_X A f \\ \cdots \\ y_{N-1} = S_X A^{N-1} f. \end{cases} \tag{4.22}$$

Similar to the work done in section 4.4, we study this problem on the Fourier domain. Due to Poisson's summation formula, we have the Lemma below.

Lemma 4.5.1. *The Fourier transform of each y_l in* (4.22) *at $(\xi,\omega) \in \mathbb{T} \times \mathbb{T}$ is*

$$\hat{y}_l(\xi,\omega) = \frac{1}{m_1 m_2} \sum_{j=0}^{m_2-1} \sum_{i=0}^{m_1-1} \hat{a}^l\Big(\frac{\xi+i}{m_1}, \frac{\omega+j}{m_2}\Big) \hat{f}\Big(\frac{\xi+i}{m_1}, \frac{\omega+j}{m_2}\Big). \tag{4.23}$$

Expression (4.23) allows for a matrix representation of the dynamical sampling problem in the case of uniform subsampling. For $j = 0,1,\cdots,m_2-1$, we define $N \times m_1$ matrices

$$A_{j,m_1,m_2}(\xi,\omega) = \left(\hat{a}^k\Big(\frac{\xi+l}{m_1}, \frac{\omega+j}{m_2}\Big) \right)_{k,l},$$

where $k = 0,1,\cdots,N-1$, $l = 0,1,\cdots,m_1-1$ and denote by $\mathscr{A}_{m_1,m_2}(\xi,\omega)$ the block matrix

$$[A_{0,m_1,m_2}(\xi,\omega)\, A_{1,m_1,m_2}(\xi,\omega) \ldots A_{m_2-1,m_1,m_2}(\xi,\omega)]. \tag{4.24}$$

Let $\bar{\mathbf{y}}(\xi,\omega) = (\hat{y}_0(\xi,\omega)\ \hat{y}_1(\xi,\omega)\ \ldots \hat{y}_{N-1}(\xi,\omega)\ )^T$ and

$$\bar{\mathbf{f}}(\xi,\omega) = \begin{pmatrix} \hat{f}(\frac{\xi}{m_1},\frac{\omega}{m_2}) \\ \hat{f}(\frac{\xi+1}{m_1},\frac{\omega}{m_2}) \\ \vdots \\ \hat{f}(\frac{\xi+m_1-1}{m_1},\frac{\omega}{m_2}) \\ \hat{f}(\frac{\xi}{m_1},\frac{\omega+1}{m_2}) \\ \vdots \\ \hat{f}(\frac{\xi+m_1-1}{m_1},\frac{\omega+1}{m_2}) \\ \vdots \\ \vdots \\ \hat{f}(\frac{\xi}{m_1},\frac{\omega+m_2-1}{m_2}) \\ \vdots \\ \hat{f}(\frac{\xi+m_1-1}{m_1},\frac{\omega+m_2-1}{m_2}) \end{pmatrix}. \tag{4.25}$$

Due to (4.23), it holds

$$\bar{\mathbf{y}}(\xi,\omega) = \frac{1}{m_1 m_2} \mathscr{A}_{m_1,m_2}(\xi,\omega)\bar{\mathbf{f}}(\xi,\omega). \tag{4.26}$$

Proposition 4.5.2. *We can recover any f from spatiotemporal samples defined in* (4.22) *if and only if $\mathscr{A}_{m_1,m_2}(\xi,\omega)$ as defined in* (4.24) *has full column rank m_1m_2 at a.e. $(\xi,\omega) \in \mathbb{T}\times\mathbb{T}$, where $\mathbb{T} = [0,1)$ under addition modulo 1. SSP is satisfied if and only if $A_{m_1,m_2}(\xi,\omega)$ is full rank for all $(\xi,\omega) \in \mathbb{T}\times\mathbb{T}$.*

By Proposition 4.5.2, we conclude that $N \geq m_1m_2$. In particular, if $N = m_1m_2$, then $\mathscr{A}_{m_1,m_2}(\xi,\omega)$ is a square matrix, we denote by $A_{m_1,m_2}(\xi,\omega)$ this square matrix.

Corollary 4.5.3. *When $N = m_1m_2$, the invertibility sampling property is equivalent to the*

condition:

$$\det A_{m_1,m_2}(\xi,\omega) \neq 0 \, \textit{for a.e.} \, (\xi,\omega) \in \mathbb{T}\times\mathbb{T}.$$

Since $A_{m_1,m_2}(\xi,\omega)$ has continuous entries, the stable sampling property is equivalent to

$$\det A_{m_1,m_2}(\xi,\omega) \neq 0 \, \textit{for all} \, (\xi,\omega) \in \mathbb{T}\times\mathbb{T}.$$

From here on we assume $N = m_1 m_2$. By its structure, $A_{m_1,m_2}(\xi,\omega)$ is a Vandermonde matrix, thus it is singular at $(\xi,\omega) \in \mathbb{T}\times\mathbb{T}$ if and only if some of its columns coincide. In case $A_{m_1,m_2}(\xi,\omega)$ is singular, no matter how many times we resample the evolved states $A^n f$, $n > N-1$, on the grid $\Omega_o = m_1\mathbb{Z}\times m_2\mathbb{Z}$, the additional data is not going to add anything new in terms of recovery and stability. In such a case we need to consider adding extra sampling locations to overcome the singularities of $A_{m_1,m_2}(\xi,\omega)$.

4.5.1 Additional sampling locations

If $A_{m_1,m_2}(\xi,\omega)$ is singular at some (ξ,ω), then by Corollary 4.5.3 the recovery of $f \in \ell^2(\mathbb{Z}^2)$ is not stable. To remove the singularities and achieve stable recovery, some extra sampling locations need to be added. The additional sampling locations depend on the positions of the singularities of $A_{m_1,m_2}(\xi,\omega)$ that we want remove. We propose a quasi-uniform way of constructing the extra sampling locations and give a characterization specifying when the singularity will be removed. Then, we use this method to remove the singularity of a *strict quadrantally symmetric* convolution operator.

Let the additional sampling set be given by

$$\Omega = \{X + (c_1,c_2) \mid (c_1,c_2) \in W \subseteq \mathbb{Z}_{m_1}\times\mathbb{Z}_{m_2}\}. \tag{4.27}$$

Let T_{c_1,c_2} denote the translation operator on $\ell^2(\mathbb{Z}^2)$, so that $T_{c_1,c_2}f(k,l) = f(k+c_1,l+c_2)$ for all $(k,l) \in \mathbb{Z}^2$. We employ a shifted sampling operator $S_X T_{c_1,c_2}$ to take extra samples at

the initial time instance; this means that our subsampling grid is shifted from $X = m_1\mathbb{Z} \times m_2\mathbb{Z}$ to $(c_1, c_2) + X$ and the extra samples are given as

$$h^{c_1,c_2}_{m_1,m_2} = S_{m_1,m_2} T_{c_1,c_2} f,\ (c_1, c_2) \in \Omega. \tag{4.28}$$

Set

$$u_{c_1,c_2}(s, p) = e^{2\pi i \frac{c_1 s}{m_1}} e^{2\pi i \frac{c_2 p}{m_2}},$$

for $(s, p) \in \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$.

By taking the Fourier transform of the samples on the additional sampling set Ω, we obtain

$$\hat{h}^{c_1,c_2}_{m_1,m_2}(\xi, \omega) = \frac{e^{2\pi i\left(\frac{c_1\xi}{m_1} + \frac{c_2\omega}{m_2}\right)}}{m_1 m_2} \sum_{s=0}^{m_1-1} \sum_{p=0}^{m_2-1} u_{c_1,c_2}(s, p) \hat{f}\left(\frac{\xi + s}{m_1}, \frac{\omega + p}{m_2}\right). \tag{4.29}$$

where

$$u_{c_1,c_2}(s, p) = e^{2\pi i \frac{c_1 s}{m_1}} e^{2\pi i \frac{c_2 p}{m_2}}.$$

For each $(c_1, c_2) \in W$, we define a row vector

$$u_{c_1,c_2} = \{u_{c_1,c_2}(s, p)\}_{(s,p)\in X}$$

with terms arranged in the same order as the terms in vector $\bar{\mathbf{f}}(\xi, \omega)$ in (4.25). We organize the vectors u_{c_1,c_2} in a matrix $\bar{U} = (u_{c_1,c_2})_{(c_1,c_2)\in W}$ and extend the data vector $\bar{\mathbf{y}}(\xi, \omega)$ in (4.26) into a big vector $\mathbf{Y}(\xi, \omega)$ by adding

$$\{e^{2\pi i \frac{-c_1\xi}{m_1}} e^{2\pi i \frac{-c_2\omega}{m_2}} (S_{m_1,m_2} T_{c_1,c_2} f)\hat{}(\xi, \omega)\}_{(c_1,c_2)\in W}.$$

Then (4.23) and (4.29) can be combined into the following matrix equation

$$\mathbf{Y}(\xi,\omega) = \frac{1}{m_1 m_2} \begin{pmatrix} \bar{U} \\ A_{m_1,m_2}(\xi,\omega) \end{pmatrix} \bar{\mathbf{f}}(\xi,\omega). \tag{4.30}$$

Proposition 4.5.4. *If a left inverse for*

$$\begin{pmatrix} \bar{U} \\ A_{m_1,m_2}(\xi,\omega) \end{pmatrix}$$

exists for every $(\xi,\omega) \in \mathbb{T}^2$*, then the vector* f *can be uniquely and stably recovered from the combined samples* (4.22) *and* (4.27) *via* (4.30)*.*

If the following property holds true:

$$\ker(\bar{U}) \cap \ker(A_{m_1,m_2}(\xi,\omega)) = 0 \tag{4.31}$$

for every (ξ,ω) in $\mathbb{T}^2$, we say that W *removes the singularities* of $A_m(\xi,\omega)$; In such a case, the assumption in Proposition 4.5.4 is satisfied.

Corollary 4.5.5. *If* W *removes the singularities of* $A_m(\xi,\omega)$ *then*

$$|W| \geq dim(\ker(A_{m_1,m_2}(\xi,\omega)))$$

for every (ξ,ω)*.*

4.5.2 Strict quadrantal symmetric convolution operator

We consider a filter a, such that $\hat{a}$ has the *strict quadrantal symmetry property*, i.e. $\hat{a}(\xi_1,\omega_1) = \hat{a}(\xi_2,\omega_2)$ for $(\xi_1,\omega_1),(\xi_2,\omega_2) \in \mathbb{T}\times\mathbb{T} = \mathbb{T}^2$ if and only if one of the following

conditions is satisfied:

$$\begin{aligned}
&1. \quad \xi_1 = \xi_2,\ \omega_1 + \omega_2 = 1 \\
&2. \quad \xi_1 + \xi_2 = 1,\ \omega_1 = \omega_2 \\
&3. \quad \xi_1 + \xi_2 = 1,\ \omega_1 + \omega_2 = 1.
\end{aligned}$$

The following result is a direct consequence of the symmetries assumptions listed in conditions $1-3$.

Proposition 4.5.6. *If $\hat{a}(\xi,\omega)$ has the strict quadrantal symmetry property, then we have* $\det A_{m_1,m_2}(\xi,\omega) = 0$ *when $\xi = 0$ or $\omega = 0$. Moreover, the kernel of each $A_{m_1,m_2}(\xi,\omega)$ is a subspace of the kernel of one of the following four matrices:*

$$A_{m_1,m_2}(0,0),\ A_{m_1,m_2}\left(\frac{1}{2},0\right),\ A_{m_1,m_2}\left(0,\frac{1}{2}\right), A_{m_1,m_2}\left(\frac{1}{2},\frac{1}{2}\right).$$

From Proposition 4.5.6, for a strict quadrantally symmetric kernel we need to consider only the points $(\xi,\omega) \in \left\{(0,0),\left(0,\frac{1}{2}\right),\left(\frac{1}{2},0\right),\left(\frac{1}{2},\frac{1}{2}\right)\right\}$ and construct the set W, such that it removes the singularities of the above four matrices.

Proposition 4.5.7. *If $\hat{a}$ has the strict quadrantal symmetry property, then*

$$\dim(A_{m_1,m_2}(\xi,\omega)) = \frac{(m_1-1)m_2}{2} + \frac{m_2-1}{2}\frac{m_1+1}{2}$$

for every $(\xi,\omega) \in \left\{(0,0),\left(0,\frac{1}{2}\right),\left(\frac{1}{2},0\right),\left(\frac{1}{2},\frac{1}{2}\right)\right\}$.

Proof. We discuss here in depth only the case $\xi = \omega = \frac{1}{2}$. The proof in the other three cases are analogous to what we present here. Because $A_{m_1,m_2}(\frac{1}{2},\frac{1}{2})$ is a Vandermonde matrix, the rank is equal to the number of its different columns. It is easy to show that

$$\hat{a}\left(\frac{\frac{1}{2}+s}{m_1},\frac{\frac{1}{2}+p}{m_2}\right) = \hat{a}\left(\frac{\frac{1}{2}+k}{m_1},\frac{\frac{1}{2}+l}{m_2}\right)$$

is satisfied if and only if one of the following holds true:

1. $s = k,\ \ p + l = m_2 - 1$

2. $p = l,\ \ s + k = m_1 - 1$

3. $s + k = m_1 - 1,\ \ p + l = m_2 - 1$

using which we can easily compute that

$$\dim(\mathscr{A}_{m_1,m_2}(\frac{1}{2},\frac{1}{2})) = \frac{(m_1-1)m_2}{2} + \frac{m_2-1}{2}\frac{m_1+1}{2} = n.$$

□

Let

$$W = W_1 \cup W_2 \tag{4.32}$$

where

$$W_1 = \{1, \cdots \frac{m_1-1}{2}\} \times \{0, \cdots, m_2 - 1\},$$

$$W_2 = \{0, \cdots, m_1 - 1\} \times \{1, \cdots, \frac{m_2-1}{2}\}\}.$$

Remark 4.5.8. When W is defined as in (4.32), we have

$$|W| = \frac{(m_1-1)m_2}{2} + \frac{m_2-1}{2}\frac{m_1+1}{2};$$

By Corollary 4.5.5, W has the minimal possible size.

Theorem 4.5.9. *Let $a \in \ell^1(D)$ be the filter such that the evolution operator is given by $Ax = a * x$. Suppose $\hat{a}$ satisfies the strict quadrantal symmetric property defined at the beginning of subsection 4.5.2. Let Ω be as in* (4.27) *with W specified in* (4.32)*. Then, any $f \in \ell^2(D)$ can be recovered in a stable way from the expanded set of samples*

$$\{S_\Omega f, S_X f, \cdots, S_X A^{m_1 m_2 - 1} f\}. \tag{4.33}$$

Proof. It suffices to show that for every $(\xi,\omega)\in\mathbb{T}\times\mathbb{T}$, it holds

$$ker(\bar{U})\cap ker(A_{m_1,m_2}(\xi,\omega))=0. \tag{4.34}$$

By Proposition 4.5.6, we only need to study the kernels of these four matrices

$$A_{m_1,m_2}(0,0), A_{m_1,m_2}(\frac{1}{2},0), A_{m_1,m_2}(0,\frac{1}{2}), A_{m_1,m_2}(\frac{1}{2},\frac{1}{2}). \tag{4.35}$$

We discuss here in depth for the case $\xi=\omega=\frac{1}{2}$. $Z:=ker(A_{m_1,m_2}(\frac{1}{2},\frac{1}{2}))$ is a subspace in $\mathbb{C}^{m_1m_2}$. By Proposition 4.5.7, the dimension of Z is n. Taking advantage of the fact that $A_{m_1,m_2}(\frac{1}{2},\frac{1}{2})$ is a Vandermonde matrix, we can choose a basis $\{v_j : j=1,\cdots,n\}$ for Z, such that each v_j has only two nonzero entries 1 and -1. Let $\mathbf{v}\in\ker(\bar{U})\cap Z$, there exists $\mathbf{c}=(c(i))_{i=1,\cdots,n}$ such that $\mathbf{v}=\sum_{i=1}^{n}c(i)v_i$. Define a $n\times n$ matrix R with the row corresponds to a fixed $(c_1,c_2)\in W$ is

$$\begin{aligned}&[(e^{\frac{2\pi i(m_1-1)c_1}{m_1}}-e^{\frac{2\pi i0c_1}{m_1}})\mathscr{F}_2(c_2),\cdots,(e^{\frac{2\pi i(m_1+1)c_1}{2m_1}}-e^{\frac{2\pi i(m_1-3)c_1}{2m_1}})\mathscr{F}_2(c_2),\\&(e^{\frac{2\pi i(m_2-1)c_2}{m_2}}-e^{\frac{2\pi i0c_2}{m_2}})\bar{\mathscr{F}}_1(c_1),\cdots,(e^{\frac{2\pi i(m_2+1)c_2}{2m_2}}-e^{\frac{2\pi i(m_2-3)c_2}{2m_2}})\bar{\mathscr{F}}_1(c_1)].\end{aligned}$$

Then

$$\bar{U}\mathbf{v}=0, \text{which is equivalent to } R\mathbf{c}=0.$$

By the use the same strategy as in the proof of Theorem 4.4.4, it can be demonstrated that these n row vectors of R are linearly independent. With slight adaptations of the strategy used so far,we can come to the same conclusion for the other three matrices in (4.35). As a consequence of Proposition 4.5.4, stability is achieved. □

4.6 Concluding Remarks

In this chapter, we have studied the spatiotemporal trade off in the two variable discrete spatially invariant evolution system driven by a single convolution filter in both finite and infinite case. We have characterized the spectral properties of the filters to recover the initial state from the uniform undersampled future states and a way to add extra spatial sampling locations to stably recover the signal when the filters violate our characterization. Compared to the one variable case, the singularity problems caused by the structure of the filters are more complicated and tougher to solve. We give explicit constructions of the extra spatial sampling locations to resolve the singularity issue caused by the strict quadrantal symmetric filters. Our results can be adapted to the general multivariable case. Different kinds of symmetry assumptions can be imposed on the filters. The problem of finding the right additional spatiotemporal sampling locations for other types of filters remains open and requires further study.

BIBLIOGRAPHY

[1] G. Pérez-Villalón A. G. García, *Multivariate generalized sampling in shift-invariant spaces and its approximation properties*, J. Math. Anal. Appl. (2009), no. 355, 397–413.

[2] Roza Aceska, Armenak Petrosyan, and Sui Tang, *Multidimensional signal recovery in discrete evolution systems via spatiotemporal trade off*, Sampl. Theory Signal Image Process. **14** (2015), no. 2, 153–169.

[3] Akram Aldroubi, Anatoly Baskakov, and Ilya Krishtal, *Slanted matrices, Banach frames, and sampling*, J. Funct. Anal. **255** (2008), no. 7, 1667–1691. MR 2442078 (2010a:46059)

[4] Akram Aldroubi, Carlos Cabrelli, Ursula Molter, and Sui Tang, *Dynamical sampling*, Appl. Comput. Harmon. Anal. (2015).

[5] Akram Aldroubi, Jacqueline Davis, and Ilya Krishtal, *Dynamical sampling: Time–space trade-off*, Appl. Comput. Harmon. Anal. **34** (2012), no. 3, 495–503.

[6] Akram Aldroubi, Jacqueline Davis, and Ilya Krishtal, *Exact reconstruction of signals in evolutionary systems via spatiotemporal trade-off*, J. Fourier Anal. Appl. **21** (2015), 11–31.

[7] Akram Aldroubi and Ilya Krishtal, *Krylov subspace methods in dynamical sampling*, Preprint, arXiv:1412.1538.

[8] Jorge Antezana, *Private communication*, (2014).

[9] R. Badeau, B. David, and G. Richard, *High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials*, IEEE Trans. Signal Process. **54** (2006), no. 4, 1341–1350.

[10] Ola Bratteli and Palle Jorgensen, *Wavelets through a looking glass*, Applied and Numerical Harmonic Analysis, Birkhäuser Boston Inc., Boston, MA, 2002, The world of the spectrum. MR 1913212 (2003i:42001)

[11] J. A. Cadzow, *Signal enhancement—a composite property mapping*, IEEE Trans. Acoust., Speech, Signal Process. **36** (1988), 49–67.

[12] Jameson Cahill, Peter G. Casazza, and Shidong Li, *Non-orthogonal fusion frames and the sparsity of fusion frame operators*, J. Fourier Anal. Appl. **18** (2012), no. 2, 287–308. MR 2898730

[13] Peter G. Casazza, Ole Christensen, Alexander M. Lindner, and Roman Vershynin, *Frames and the Feichtinger conjecture*, Proc. Amer. Math. Soc. **133** (2005), no. 4, 1025–1033 (electronic). MR 2117203 (2006a:46024)

[14] Peter G. Casazza, Gitta Kutyniok, and Shidong Li, *Fusion frames and distributed processing*, Appl. Comput. Harmon. Anal. **25** (2008), no. 1, 114–132. MR 2419707 (2009d:42094)

[15] Peter G. Casazza and Janet Crandell Tremain, *The Kadison-Singer problem in mathematics and engineering*, Proc. Natl. Acad. Sci. USA **103** (2006), no. 7, 2032–2039 (electronic). MR 2204073 (2006j:46074)

[16] John B. Conway, *A course in functional analysis*, Graduate Texts in Mathematics, vol. 96, Springer-Verlag, New York, 1985. MR 768926 (86h:46001)

[17] Bradley Currey and Azita Mayeli, *Gabor fields and wavelet sets for the Heisenberg group*, Monatsh. Math. **162** (2011), no. 2, 119–142. MR 2769882 (2012d:42069)

[18] Ingrid Daubechies, *Ten lectures on wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. MR 1162107 (93e:42045)

[19] A. Eisinberg, G. Franzé, and P. Pugliese, *Vandermonde matrices on chebyshev points*, Linear Algebra and its Applications **283** (1998), no. 1-3, 205–219.

[20] Brendan Farrell and Thomas Strohmer, *Inverse-closedness of a Banach algebra of integral operators on the Heisenberg group*, J. Operator Theory **64** (2010), no. 1, 189–205. MR 2669435

[21] W. Gautschi, *Norm estimates for inverses of vandermonde matrices*, Numer. Math. **23** (1975), 337–347.

[22] A. Gilbert, P. Indyk, M. Iwen, and L. Schmidt, *Recent developments in the sparse fourier transform*, IEEE Signal Processing Magazine **31** (2014), no. 5, 91–100.

[23] Karlheinz Gröchenig, *Localization of frames, Banach frames, and the invertibility of the frame operator*, J. Fourier Anal. Appl. **10** (2004), no. 2, 105–132. MR 2054304 (2005f:42086)

[24] Karlheinz Gröchenig and Michael Leinert, *Wiener's lemma for twisted convolution and Gabor frames*, J. Amer. Math. Soc. **17** (2004), no. 1, 1–18 (electronic). MR 2015328 (2004m:42037)

[25] Deguang Han and David Larson, *Frame duality properties for projective unitary representations*, Bull. Lond. Math. Soc. **40** (2008), no. 4, 685–695. MR 2441141 (2009g:42057)

[26] Christopher Heil, *A basis theory primer*, expanded ed., Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2011. MR 2744776 (2012b:46022)

[27] Eugenio Hernández and Guido Weiss, *A first course on wavelets*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1996, With a foreword by Yves Meyer. MR 1408902 (97i:42015)

[28] Kenneth Hoffman, *Banach spaces of analytic functions*, Dover Publications, Inc., New York, 1988, Reprint of the 1962 original. MR 1102893 (92d:46066)

[29] Kenneth Hoffman and Ray Kunze, *Linear algebra*, Second edition, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. MR 0276251 (43 #1998)

[30] A. Hormati, O. Roy, Y.M. Lu, and M. Vetterli, *Distributed sampling of signals linked by sparse filtering: Theory and applications*, Signal Processing, IEEE Transactions on **58** (2010), no. 3, 1095 –1109.

[31] Yingbo Hua and T. K. Sarar, *Matrix pencil method for estimating parameters of exponentially damped/ undamped sinusoids in noise*, IEEE Trans. Acoust., Speech, Signal Process. **38** (1990), 814–824.

[32] ———, *On svd for estimating generalized eigenvalues of singular matrix pencil in noise*, IEEE Trans. Signal Process. **39** (1991), no. 4, 892–900.

[33] M. Iwen, *Combinatorial sublinear-time fourier algorithms*, Foundations of Computational Mathematics **10** (2010), 303–338.

[34] ———, *Improved approximation guarantees for sublinear-time fourier algorithms*, Applied And Computational Harmonic Analysis **34** (2013), 57–82.

[35] Y.M. Lu and M. Vetterli, *Spatial super-resolution of a diffusion field by temporal oversampling in sensor networks*, Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, april 2009, pp. 2249–2252.

[36] Stéphane Mallat, *A wavelet tour of signal processing*, Academic Press Inc., San Diego, CA, 1998. MR 1614527 (99m:94012)

[37] A. Marcus, D. Spielman, and N. Srivastave, *Interlacing families ii: Mixed characteristic polynomials and the kadison–singer problem*, Annals of Mathematics **182** (2015), no. 1, 327–350.

[38] M. Zuhair Nashed, *Inverse problems, moment problems, signal processing: un menage a trois*, Mathematics in science and technology, World Sci. Publ., Hackensack, NJ, 2011, pp. 2–19. MR 2883419

[39] Gestur Ólafsson and Darrin Speegle, *Wavelets, wavelet sets, and linear actions on* $\mathbb{R}^n$, Wavelets, frames and operator theory, Contemp. Math., vol. 345, Amer. Math. Soc., Providence, RI, 2004, pp. 253–281. MR 2066833 (2005h:42075)

[40] Thomas Peter and Gerlind Plonka, *A generalized prony method for reconstruction of sparse sums of eigenfunctions of linear operators*, Inverse Problems **29** (2013), 025001.

[41] R. Roy, A. Paulraj, and T. Kailath, *Esprit-a subspace rotation ap- proach to estimation of parameters of cisoids in noise*, IEEE Trans. Acoust., Speech, Signal Process. **34** (1986), no. 5, 1340–1342.

[42] D. Slepian and J. Wolf, *Noiseless coding of correlated information sources*, Information Theory, IEEE Transactions on **19** (1973), no. 4, 471 – 480.

[43] Gilbert Strang and Truong Nguyen, *Wavelets and filter banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996. MR 1411910 (98b:94003)

[44] Qiyu Sun, *Frames in spaces with finite rate of innovation*, Adv. Comput. Math. **28** (2008), no. 4, 301–329. MR 2390281 (2009c:42093)

[45] Sui Tang, *System identification in dynamical sampling*, Preprint, arxiv:1502.02741v2, 2015.

[46] A. Timan, *Theory of approximation of functions of a real variable*, Courier Dover Publications, page 37 1963.

[47] E. B. Vinberg, *A course in algebra*, ISBN 0-8218-3413-4, American Mathematical Society, Providence, R.I, 2003.