SUBSPACE APPROXIMATION ON THE CONTINUUM

by

Zhihui Zhu

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Electrical Engineering).

Golden, Colorado

Date _____

Signed: _____
Zhihui Zhu

Signed: _____
Dr. Michael B. Wakin
Thesis Advisor

Golden, Colorado

Date _____

Signed: _____
Dr. Atef Elsherbeni
Professor, Dobelman Chair, and Department Head
Department of Electrical Engineering and Computer Science

ABSTRACT

Many signal processing problems—such as analysis, compression, reconstruction, and denoising—can be facilitated by exploiting the underlying models the signals and data sets obey. A model often deals with the notion of conciseness and suggests a signal has few degrees of freedom relative to its ambient dimensionality. For instance, the Shannon-Nyquist sampling theorem works on bandlimited signals obeying *subspace model*. As an another example, the power of sparse signal processing often relies on the assumption that the signals live in some union of subspaces. In many cases, signals have concise representations which are often obtained by (i) constructing a *dictionary* of elements drawn from the signal space, and then (ii) expressing the signal of interest as a linear combination of a small number of atoms drawn from the dictionary. Such representations serve as an efficient way to describe the conciseness of the signals and enable effective signal processing methods. For example, the sparse representation forms the core of compressive sensing (CS), an emerging research area that aims to break through the Shannon-Nyquist limit for sampling analog signals.

However, despite its recent success, there are many important applications in signal processing that do not naturally fall into the subspace models and sparse recovery framework. As a classical example, a finite-length vector obtained by sampling a bandlimited signal is not sparse using the discrete Fourier transform (DFT), the natural tool for frequency analysis on finite-dimensional space. In other words, the DFT cannot excavate the concise structure within the sampled bandlimited signals. These signals obey a so-called *parameterized subspace model* in which the signals of interest are inherently low-dimensional and live in a union of subspaces, but the choice of subspace is controlled by a small number of continuous-valued parameters (the parameter controlling sampled bandlimited signals is the frequency). This continuous-valued parameterized subspace model appears in many problems including spectral estimation, mitigation of narrowband interference, feature extraction, and steerable

filters for rotation-invariant image recognition.

The purpose of this thesis is to 1) construct a subspace—whose dimension matches the effective number of local degrees of freedom—for approximating (almost) all the signals controlled by a small number of continuous-valued parameters ranging within some certain intervals; 2) develop rigorous, theoretically-backed techniques for computing projections onto and orthogonal to these subspaces. By developing an appropriate basis to economically represent the signals of interest, one can apply effective tools developed for the subspace model and sparse recovery framework for signal processing. In the process of building local subspace fits, we will also obtain the effective dimensionality of such signals.

Our key contributions include (i) new non-asymptotic results on the eigenvalue distribution of (periodic) discrete time-frequency localization operators and fast constructions for computing approximate projections onto the *discrete prolate spheroidal sequences* (DPSS's) subspace; (ii) an orthogonal approximate Slepian transform that has computational complexity comparable to the fast Fourier transform (FFT); (iii) results on the spectrum of combined time- and multiband-limiting operations in the discrete-time domain and analysis for a dictionary formed by concatenating a collection of modulated DPSS's; (iv) analysis for the dimensionality of wall and target return subspaces in through-the-wall radar imaging and algorithms for mitigating wall clutter and identifying non-point targets; (v) asymptotic performance guarantee of the individual eigenvalue estimates for Toeplitz matrices by circulant matrices; and (vi) analysis of the eigenvalue distribution of time-frequency limiting operators on locally compact abelian groups.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGMENTS

The best part of graduate school at Mines has been the chance to meet and work with so many amazing people. I have been fortunate to work with a large and talented group of collaborators: Sandeep Gogineni, Qiuwei Li, Shuang Li, Hassan Mansour, Muralidhar Rangaswamy, Pawan Setlur, Gongguo Tang, Xinming Wu, Dehui Yang, and especially Mark Davenport, Armin Eftekhari, Santhosh Karnik, Justin Romberg, and Michael Wakin with whom I collaborate closely on the project—subspace approximation on the continuum (also the title of this thesis).

I wish to express my gratitude to my committee for their valuable suggestions and contributions: to Paul Constantine for some very challenging but motivating discussions and for organizing the enjoyable AMS colloquium; to Mark Davenport for the time and energy he generously poured into the project mentioned above and for hosting me at GaTech; to Gongguo Tang for his always helpful discussion in every technical question and for becoming such an encouraging collaborator; to Tyron Vincent for his terrific course on Estimation theory and Kalman filtering; and most of all to my advisor Michael Wakin for his extreme patience with me. I would not be the researcher I am today were it not for the countless hours Mike enthusiastically devoted to help with speaking, writing, and researching.

I also want to thank all the fun people to work with over the years at Mines: Armin, Bernard, Dehui, Gongguo, Hossein, Kai, Louis, Mike, Mohammad, Qiuwei and Shuang (the Condor Heroes), Paul, Pang, Rong, Tong, Tyron, Youye, Xinming and Xu.

Finally, I want to thank all of my other friends and family for their continued encouragement and support: Grandpa, Mom, Dad, Brother, LG (the "god father" who inspired and continuously encourages me to pursue a career in research), Chong, Feng, Tao, Xiao and everyone else who helped me along the way.

CHAPTER 1

INTRODUCTION

In this chapter, we present the central theme of this dissertation and outline our specific contributions.

1.1 Concise Signal Models

A fundamental challenge in signal processing is to efficiently acquire and extract information for (potentially high-dimensional) signals and data sets. Effective techniques for overcoming this challenge often rely on accurate *models* for signals or data sets of interest. In general, models are useful as a priori knowledge for analysis, capture, compression and storage, communication, denoising and processing of signals.

For a given problem, a model can always be specified for the signals to be processed. A model can take the form of deterministic class (like a bandlimited signal), or probabilistic model (like a stochastic process). We can utilize this model to distinguish classes of interesting signals from uninteresting ones, or information from noise. For example, based on the assumption that a continuous-time signal is bandlimited (i.e., it can be written as a linear combination of low-frequency sinusoids), the Shannon-Nyquist sampling theorem [114] specifies a minimal sampling rate that permits a discrete sequence of its *samples* to preserve all the information. At its core, the Shannon-Nyquist sampling theorem states that we can utilize the bandlimitedness of the given signal to distinguish it from other signals only from its samples.

Modelling most physical signals as bandlimited is a stylized example of *subspace model* in signal processing. Subspace models have been incorporated into modern signal processing by modeling signals as *vectors* living in an appropriate *vector space* (also called *linear space*). A vector space is a collection of vectors, which may be added together or multiplied by *scalars* resulting in new vectors that still live in this vector space. Simple but widely encountered

vector spaces include $\mathbb{R}^N$, a coordinate space with N real variables, and L_2, the set of all square-integrable functions. A linear subspace $\mathcal{S}$ of a vector space $\mathcal{V}$ is a nonempty subset of $\mathcal{V}$ that is closed under addition and scalar multiplication. A subspace is well characterized by its *dimension* and *basis*, which is a set of vectors that are linear independent and span the whole subspace. The dimensionality serves as a way to describe the conciseness of signals living in this subspace. Based on the fact that any signal in the subspace $\mathcal{S}$ can be represented as a linear combination of the basis elements (which are also called atoms), subspace modeling provides a powerful tool for signal processing. For example, one can compress the signals by using the representation coefficients. As an another example, we can denoise the signals (if they are corrupted by noise) by projecting them onto the subspace.

Like the assumption of bandlimitedness, models in signal processing often rely on the assumption that the signals have few degrees of freedom relative to their ambient dimensionality. For example, we say that a signal is *sparse* if it can be well-approximated as a linear combination of a small number of atoms or elements from some basis or *dictionary*. Sparsity captures the concise structure existing in most natural signals. As an example, many natural images have a sparse representation in a wavelet dictionary because very few wavelets can describe large smooth image regions.

Sparse representation has been widely used for signal denoising [39], signal recovery [13] and compressive sensing (CS) [15, 16, 18, 26, 40], an emerging research area that aims to break through the Shannon-Nyquist limit for sampling analog signals. One challenge in sparse modeling—due to its high nonlinearity since the choice of which dictionary elements are used can change from signal to signal—is to identify which of the dictionary elements best representing a given signal. This problem has garnered much attention in the applied mathematics and signal processing communities, and conditions can be established under which methods based on convex optimization [15, 24, 41] and greedy algorithms [11, 93, 99, 127] provide suitable approximations.

For a given class of signals or data sets, another challenge in sparse modeling is to construct a dictionary in which the set of signals have a sparse representation. This challenge is usually referred to as dictionary learning. Typical algorithms for dictionary learning with alternating minimization scheme include the method of optimal directions (MOD) [48], the K-singular value decompostion (K-SVD) [1] and a method for designing an incoherent sparsifying dictionary [87]. Though these algorithms have been known to work pretty well in practice, it remains an active research area to fully understand the underpinning of this phenomenon, especially for the conditions under which the dictionary can be exactly recovered. Recently, Sun et al. [123] provided a theoretical understanding of as well as a provable efficient algorithm for dictionary learning with a complete dictionary, i.e., the dictionary is square and is also full rank.

1.2 Parameterized Subspace Models

As we discussed above, models play a central role in signal processing. Moving to the richer subspace model has led to powerful new techniques for dimensionality reduction, noise removal, and recovering high-dimensional signals from indirect observations. However, despite its recent success, there are many important applications in signal processing that do not naturally fall into the subspace models and sparse recovery framework. We describe several representative examples below.

Scenario #1. Consider an analog signal $x(t)$ that can be expressed as a sum of just K sinusoids of various frequencies, amplitudes, and phases. Suppose we collect N Nyquist-rate samples of $x(t)$ and stack these into a sample vector $\boldsymbol{x} \in \mathbb{C}^N$. If we analyze $\boldsymbol{x}$ using the natural tool for frequency analysis on $\mathbb{C}^N$—the DFT—then in general all N DFT coefficients will be nonzero, unless the analog frequencies contributing to $x(t)$ happen to live on a certain harmonic grid. This is even though the actual number of degrees of freedom in $x(t)$ (and thus in $\boldsymbol{x}$) is just K. In other words, the DFT cannot reveal the concise structure within $\boldsymbol{x}$. Figure 1.1 severs as an example showing the DFT of a sampled pure sinusoid with frequency off the "DFT-grid". In a more practical scenario, suppose the original analog signal $x(t)$ has

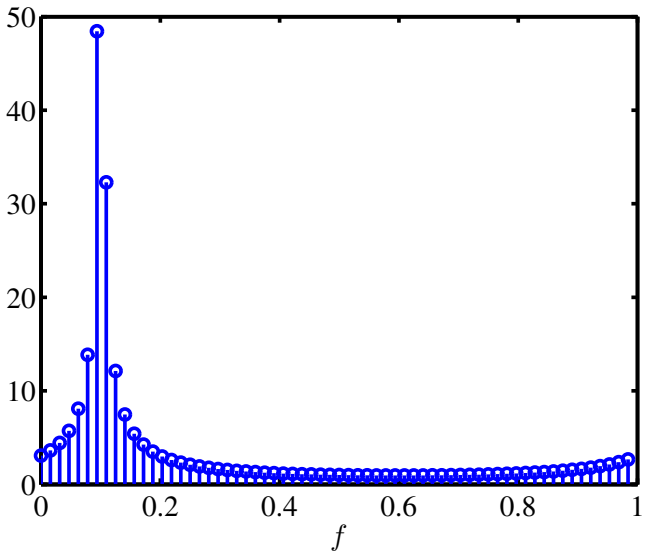


Figure 1.1: A plot of the DFT of a length-N (with $N = 64$) pure sinusoid with frequency 0.1 which is off the "DFT-grid".

a Fourier transform with support limited to a small number of narrow bands (rather than consisting of pure tones). Again, in general, all N DFT coefficients of $\boldsymbol{x}$ will be nonzero since the time-limiting operation will cause the spectral bands to smear out. We describe this scenario in Section 2.4 with more detail. This complication arises in problems such as spectral estimation and mitigation of narrowband interference.

Scenario #2. Another example arises in problems such as time of arrival estimation in matched filtering, radar signal processing with point targets, and super-resolution. Consider an analog signal $x(t)$ that can be expressed as a sum of K pulses $p(t)$ of a known shape but with differing amplitudes and times of arrival. Suppose we collect N observations of $x(t)$ and stack these into a sample vector $\boldsymbol{x} \in \mathbb{C}^N$. The observations could consist of N Nyquist-rate samples of $x(t)$, the N lowest frequency Fourier series coefficients of $x(t)$, or the inner products between $x(t)$ and N preselected linear functionals. In general there is no orthonormal basis for $\mathbb{C}^N$ in which x will be K-sparse, even though the actual number of degrees of freedom in $x(t)$ (and thus in x) is just K. In the case of non-point targets, then each of the K pulses in the original analog signal $x(t)$ is convolved with a different unknown

kernel of narrow—but nonzero—width. Again there is no basis in which the vector $\boldsymbol{x}$ can be economically represented and processed.

In the representative problems we list above, the signals obey a so-called *parameterized subspace model* in which the signals of interest are inherently low-dimensional and live in a union of subspaces, but the choice of subspace is controlled by a small number of continuous-valued parameters. This continuous-valued parameterized subspace model also appears in many other problems, such as feature extraction aiming to detect a pattern independent of its orientation in an image [86], steerable filters for rotation-invariant image recognition [133], and so on.

1.3 Overview and Contributions

This dissertation mainly focuses on parameterized subspace models. Rather than attempt to discretize these problems and use existing sparse processing techniques (a program potentially fraught with difficulty, as the sparsity generally does not translate directly into the discrete domain), our essential research goal is to develop general techniques for sparse processing using a more natural parameterized subspaces model. In particular, we aim to 1) construct a subspace for approximating (almost) all the signals controlled by a small number of continuous-valued parameters ranging within some certain intervals; 2) develop rigorous, theoretically-backed techniques for computing projections onto and orthogonal to these subspaces. By developing an appropriate basis to economically represent the signals of interest, one can apply effective tools developed for subspace modeling and sparse modeling for signal processing. In the processing of building local subspace fits, we will also provide an answer to the effective dimensionality of such signals.

Our key contributions include new:

- *non-asymptotic results* on the eigenvalue distribution of (periodic) discrete time-frequency localization operators and *fast constructions* for computing approximate projections onto the leading Slepian basis elements;

- a rapid *orthogonal* approximate Slepian transform (ROAST) for the discrete vector one obtains when collecting a finite set of uniform samples from a baseband analog signal;

- *results* on the spectrum of combined time- and multiband-limiting operations in the discrete-time domain and *analysis* for a dictionary formed by concatenating a collection of modulated DPSS's;

- *analysis* for the dimensionality of wall and target return subspaces in through-the-wall radar imaging and *algorithms* for mitigating wall clutter and identifying non-point targets;

- *asymptotic performance guarantees* of the individual eigenvalue estimates for Toeplitz matrices by circulant matrices;

- *generalization* of the existing results on the eigenvalues of composite time- and band-limiting operators to locally compact abelian groups.

We outline these contributions chapter-by-chapter.

We begin in Chapter 2 with a review of basic topics in signal processing and mathematics that form the foundation of this dissertation.

In Chapter 3, we consider the task of concisely representing a discrete vector one obtains when collecting a finite set of uniform samples from a baseband analog signal (which is a special case of **Scenario #1**). The optimal basis (i.e., solution) to this problem in the least-squares sense forms the foundation for the analysis of the problems listed in **Scenario #1** and is given by the DPSS basis (a.k.a the *Slepian basis*) [118]. However, due to the high computational complexity of projecting onto the DPSS basis, this representation is often overlooked in favor of the FFT. We show that there exist fast constructions for computing approximate projections onto the leading Slepian basis elements. The complexity of the resulting algorithms is comparable to the FFT, and scales favorably as the quality of the desired approximation is increased. In the process of bounding the complexity of these

algorithms, we also establish a new nonasymptotic result on the eigenvalue distribution of discrete time-frequency localization operators. This nonasymptotic result is also extended for the eigenvalue distribution of periodic discrete time-frequency localization operators [57, 71]. We then demonstrate how these algorithms allow us to efficiently compute the solution to certain least-squares problems that arise in signal processing. We also provide simulations comparing these fast, approximate Slepian methods to exact Slepian methods as well as traditional FFT based methods.

In Chapter 4, we provide a Rapid Orthogonal Approximate Slepian Transform (ROAST) for the discrete vector one obtains when collecting a finite set of uniform samples from a baseband analog signal. Unlike the fast construction of projections onto the leading DPSS vectors that is not an orthogonal projection, the ROAST offers an orthogonal projection which is an approximation to the orthogonal projection onto the leading DPSS vectors. As such, the ROAST is guaranteed to accurately and compactly represent not only oversampled bandlimited signals but also the leading DPSS vectors themselves. Moreover, the subspace angle between the ROAST subspace and the corresponding DPSS subspace can be made arbitrarily small. The complexity of computing the representation of a signal using the ROAST is comparable to the FFT, which is much less than the complexity of using the DPSS basis vectors. We also give non-asymptotic results to guarantee that the proposed basis not only provides a very high degree of approximation accuracy in a mean-square error sense for bandlimited sample vectors, but also that it can provide high-quality approximations of all sampled sinusoids within the band of interest.

In Chapter 5, we study possible dictionaries for representing the discrete vector one obtains when collecting a finite set of uniform samples from a multiband analog signal (the problems presented in **Scenario #1**). By analyzing the spectrum of combined discrete time- and multiband-limiting operations (which are equivalent to some corresponding Toeplitz matrices), we conclude that the information level of the sampled multiband vectors is essentially equal to the time-frequency area. For representing these vectors, we consider a dictionary

formed by concatenating a collection of modulated Discrete Prolate Spheroidal Sequences (DPSS's). We study the angle between the subspaces spanned by this dictionary and an optimal dictionary, and we conclude that the multiband modulated DPSS dictionary—which is simple to construct and more flexible than the optimal dictionary in practical applications—is nearly optimal for representing multiband sample vectors. We also show that the multiband modulated DPSS dictionary not only provides a very high degree of approximation accuracy in a mean squared error (MSE) sense for multiband sample vectors (using a number of atoms comparable to the information level), but also that it can provide high-quality approximations of all sampled sinusoids within the bands of interest.

In Chapter 6, motivated by the fact that Toeplitz matrices appear naturally for parameterized subspace models, we study a fast way to approximately compute the spectrum of Toeplitz matrices. It is known that any sequence of uniformly bounded $N \times N$ Hermitian Toeplitz matrices $\{\boldsymbol{H}_N\}$ is asymptotically equivalent to a certain sequence of $N \times N$ circulant matrices $\{\boldsymbol{C}_N\}$ derived from the Toeplitz matrices in the sense that $\|\boldsymbol{H}_N - \boldsymbol{C}_N\|_F = o(\sqrt{N})$ as $N \to \infty$. This implies that certain collective behaviors of the eigenvalues of each Toeplitz matrix are reflected in those of the corresponding circulant matrix and supports the utilization of the computationally efficient fast Fourier transform (instead of the Karhunen-Loève transform) in applications like coding and filtering. We study the asymptotic performance of the *individual eigenvalue estimates*. We show that the asymptotic equivalence of the circulant and Toeplitz matrices implies the individual asymptotic convergence of the eigenvalues for certain types of Toeplitz matrices. We also show that these estimates asymptotically approximate the largest and smallest eigenvalues for more general classes of Toeplitz matrices.

In Chapter 7, we consider the eigenvalues of composite time- and band-limiting operators on locally compact abelian groups. By invoking Fourier transforms for functions defined on locally compact abelina groups, the time-frequency limiting operators generalize the conventional limiting operators which result in DPSS's and *prolate spheroidal wave functions* (PSWF's) [120]. Applications of this unifying treatment are discussed in relation to channel

capacity and to representation and approximation of signals.

We conclude with a final discussion and directions for future research in Chapter 8.

This thesis is a reflection of a series of intensive collaborations. Where appropriate, the first page of each chapter lists primary collaborators, who share credit for this work.

CHAPTER 2

BACKGROUND MATERIAL

In this chapter, we briefly review basic topics in signal processing and linear algebra.

2.1 General Mathematical Preliminaries

To begin, we provide a brief discussion of mathematical preliminaries and notation. We indicate finite-dimensional vectors and matrices by bold characters to distinguish them from infinite-length sequences and signals (or functions) with continuous variables. We index all such vectors and matrices beginning at 0. The n-th element of a vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}[n]$, while the (m,n)-th element of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}[m,n]$. The Hermitian transpose of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^*$ or $\boldsymbol{A}^{\mathrm{H}}$. For any natural number N, we let $[N]$ denote the set $\{0,1,\ldots,N-1\}$. For any $k \in \{1,2,\ldots,N\}$, let $[\boldsymbol{A}]_k$ denote the $N \times k$ matrix formed by taking the first k columns of $\boldsymbol{A} \in \mathbb{C}^{N\times N}$. We use $o(\cdot)$ and $O(\cdot)$ as the conventional "little-o" and "big-O" notations, respectively. In addition, $x(N) \sim y(N)$ means x and y are asymptotically equal, that is $x(N) = y(N) + o(y(N)) = (1+o(1))y(N)$ as $N \to \infty$. For any $f \in [-\frac{1}{2}, \frac{1}{2}]$ we will let

$$\boldsymbol{e}_f := \begin{bmatrix} e^{j2\pi f 0} \\ e^{j2\pi f 1} \\ \vdots \\ e^{j2\pi f(N-1)} \end{bmatrix} \in \mathbb{C}^N \tag{2.1}$$

denote a length-N vector of samples from a discrete-time complex exponential signal with digital frequency f. This sampled exponential $\boldsymbol{e}_f$ along with its corresponding $N \times N$ matrix $\boldsymbol{E}_f := \mathrm{diag}(\boldsymbol{e}_f)$ will appear very often through this dissertation. Here, $\mathrm{diag} : \mathbb{C}^N \to \mathbb{C}^{N\times N}$ returns a square diagonal matrix with the elements of the input vector on the main diagonal.

For a given continuous-time function $x(t), t \in \mathcal{D}$, the conventional L_p norm with $p \geq 1$ is defined as

$$\|x\|_{L_p(\mathcal{D})} = \left(\int_{t \in \mathcal{D}} |x(t)|^p \, \mathrm{d}\, t \right)^{1/p}.$$

An $L_p(\mathcal{D})$ space consists of all functions for which the L_p norm is finite (i.e., the p-th power of the absolute value is integrable). When $\mathcal{D} = \mathbb{R}$, we usually drop it and just write L_p space and $\|x\|_{L_p}$. In $\mathbb{R}^N$ or $\mathbb{C}^N$, the p-norm of $\boldsymbol{x} \in \mathbb{C}^N$ is defined by

$$\|\boldsymbol{x}\|_p = \left(\sum_{n=0}^{N-1} |\boldsymbol{x}[n]|^p \right)^{1/p}.$$

For any infinite-length sequence x, its p-norm is also defined as

$$\|x\|_p = \left(\sum_{n=-\infty}^{\infty} |x[n]|^p \right)^{1/p}.$$

An $\ell_p(\boldsymbol{Z})$ space consists of all sequences for which the ℓ_p norm is finite.

2.2 Conventional Fourier Transforms

For a given continuous-time function $x(t) \in L_1$, its *continuous-time Fourier transform (CTFT)* is defined by

$$X(F) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi F t} \, \mathrm{d}\, t,$$

where the transform variable F (usually) represents *frequency* when the independent variable t represents *time*. The corresponding *inverse CTFT (ICTFT)* is defined as

$$x(t) = \int_{-\infty}^{\infty} X(F) e^{j2\pi F t} \, \mathrm{d}\, F.$$

For any $\boldsymbol{x} \in \mathbb{C}^N$, the *discrete Fourier transform (DFT)* of $\boldsymbol{x}$, denoted by $\widehat{\boldsymbol{x}}$, is defined as

$$\widehat{\boldsymbol{x}}[n] := \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \boldsymbol{x}[n] e^{-j\frac{2\pi n m}{N}}$$

for all $n \in [N]$. Given $\widehat{\boldsymbol{x}}$, $\boldsymbol{x}$ can be recovered by taking the *inverse DFT (IDFT)*, i.e.,

$$\boldsymbol{x}[m] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \widehat{\boldsymbol{x}}[n] e^{j\frac{2\pi n m}{N}}.$$

For a given discrete-time signal $x[n]$, we let

$$\widetilde{x}(f) = \sum_{n=-\infty}^{\infty} x[n]e^{-j2\pi fn}$$

denote the *discrete-time Fourier transform* (DTFT) of $x[n]$. The corresponding *inverse DFTF (IDTFT)* is defined as

$$x[n] = \int_0^1 \widetilde{x}(f)e^{j2\pi fn}$$

for all $n \in \mathbb{Z}$. For any $\boldsymbol{x} \in \mathbb{C}^N$, we use $\widehat{\boldsymbol{x}}$ to denote the DTFT of the sequence obtained by zero-padding $\widehat{\boldsymbol{x}}$, i.e.

$$\widetilde{\boldsymbol{x}}(f) = \sum_{n=0}^{N} x[n]e^{-j2\pi fn}.$$

2.3 Signal Dictionaries and Representations

Effective techniques for signal processing often rely on meaningful representations that capture the structure inherent in the signals of interest. Many signal processing tasks—such as signal denoising, recognition, and compression—benefit from having a *concise* signal representation. Concise signal representations are often obtained by (i) constructing a *dictionary* of elements drawn from the signal space, and then (ii) expressing the signal of interest as a linear combination of a small number of atoms drawn from the dictionary.

For the signal space $\mathbb{C}^N$, we represent a dictionary as an $N \times L$ matrix $\boldsymbol{\Psi}$, which has columns (or atoms) $\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{L-1}$. Using this dictionary, a signal $\boldsymbol{x} \in \mathbb{C}^N$ can be represented exactly or approximately as a linear combination of the $\boldsymbol{\psi}_i$:

$$\boldsymbol{x} \approx \boldsymbol{\Psi}\boldsymbol{\alpha} = \sum_{i=0}^{L-1} \boldsymbol{\alpha}[i]\boldsymbol{\psi}_i$$

for some $\boldsymbol{\alpha} \in \mathbb{C}^L$, whose entries are referred to as coefficients.

When the coefficients have a small fraction of nonzero values or decay quickly, one can form highly accurate and concise approximations of the original signal using just a small number of atoms. In some cases, one can achieve this using a linear approximation that is

formed with a prescribed subset of $J < L$ atoms:

$$\boldsymbol{x} \approx \sum_{i \in \Omega} \boldsymbol{\alpha}[i] \boldsymbol{\psi}_i, \tag{2.2}$$

where $\Omega \subset \{0, 1, \ldots, L-1\}$ is a fixed subset of cardinality J. For example, one might use the lowest J frequencies to approximate bandlimited signals in a Fourier basis.

In other cases, it may be beneficial to adaptively choose a set of atoms in order to optimally represent each signal. Such a nonlinear approximation can be expressed as

$$\boldsymbol{x} \approx \sum_{i \in \Omega(\boldsymbol{x})} \boldsymbol{\alpha}[i] \boldsymbol{\psi}_i,$$

where $\Omega(\boldsymbol{x}) \subset \{0, 1, \ldots, L-1\}$ is a particular subset of cardinality J and can change from signal to signal. A more thorough discussion of this topic, which is also known as *sparse approximation*, can be found in [34, 37, 92]. Sparse approximations have been widely used for signal denoising [39], signal recovery [13] and compressive sensing (CS) [15, 16, 18, 26, 40], an emerging research area that aims to break through the Shannon-Nyquist limit for sampling analog signals. A challenge in finding the best J-term approximation for a given signal $\boldsymbol{x}$ is to identify which of the $\binom{L}{J}$ subspaces (or, equivalently, index sets $\Omega(\boldsymbol{x})$) to use. Many methods based on convex optimization [15, 24, 41] and greedy algorithms [11, 93, 99, 127] provide suitable approximations under certain conditions.

2.4 Finite-length Vectors of Sampled Analog Signals

As a motivating example listed in Section 1.2, we will study dictionaries for representing the discrete vector one obtains when collecting a finite set of uniform samples from a certain type of analog signal. We let $x(t)$ denote a complex-valued analog (continuous-time) signal, and for some finite number of samples N and some sampling period $T_s > 0$, we let

$$\boldsymbol{x} = [x(0) \; x(T_s) \; \cdots \; x((N-1)T_s)]^T \tag{2.3}$$

denote the length-N vector obtained by uniformly sampling $x(t)$ over the time interval $[0, NT_s)$ with sampling period T_s. Here T stands for the transpose operator. Our focus is on obtaining a dictionary $\boldsymbol{\Psi}$ that provides highly accurate approximations of $\boldsymbol{x}$ using as few atoms as possible.

It is the structure we assume in the analog signal $x(t)$ that motivates the search for a concise representation of $\boldsymbol{x}$. Specifically, we assume that $x(t)$ obeys a multiband signal model, in which the signal's continuous-time Fourier transform (CTFT) is supported on a small number of narrow bands (we assume the bands are known). We describe this model more fully in Section 2.4.2. Before doing so, we begin in Section 2.4.1 with a simpler analog signal model for which an efficient dictionary $\boldsymbol{\Psi}$ is easier to describe.

2.4.1 Multitone Signals

A *multitone* analog signal is one that can be expressed as a sum of J complex exponentials of various frequencies:

$$x(t) = \sum_{i=0}^{J-1} \beta_i e^{j2\pi F_i t}.$$

Suppose such a multitone signal $x(t)$ is bandlimited with bandlimit $\frac{B_{\text{nyq}}}{2}$ Hz, i.e., that $\max_i |F_i| \leq \frac{B_{\text{nyq}}}{2}$. Let $\boldsymbol{x}$, as defined in (2.3), denote the length-N vector obtained by uniformly sampling $x(t)$ over the time interval $[0, NT_s)$ with sampling period $T_s \leq \frac{1}{B_{\text{nyq}}}$ which meets the Nyquist sampling rate. We can express these samples as

$$\boldsymbol{x}[n] = \sum_{i=0}^{J-1} \beta_i e^{j2\pi f_i n}, \ n = 0, 1, \ldots, N-1, \tag{2.4}$$

where $f_i = F_i T_s$. This model arises in problems such as radar signal processing with point targets [78] and super-resolution [17].

In certain cases, an effective dictionary for representing $\boldsymbol{x}$ is the $N \times N$ DFT matrix [7, 128, 78], where $\boldsymbol{\psi}_i[n] = e^{j2\pi in/N}$ for $i = 0, 1, \ldots, N-1$ and $n = 0, 1, \ldots, N-1$. Using this dictionary, we can write $\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{C}^N$ contains the DFT coefficients of $\boldsymbol{x}$. When the frequencies f_i appearing in (2.4) are all integer multiples of $1/N$, then $\boldsymbol{\alpha}$ will be

J-sparse (meaning that it has at most J nonzero entries), and the sparse structure of $x(t)$ in the analog domain will directly translate into a concise representation for $\boldsymbol{x}$ in $\mathbb{C}^N$. This "on grid" multitone signal is sometimes assumed for simplicity in the CS literature [128]. However, when the frequencies comprising $x(t)$ are arbitrary, the sparse structure in $\boldsymbol{\alpha}$ will be destroyed due to the "DFT leakage" phenomenon. Such a problem can be mitigated by applying a windowing function in the sampling system, as in [128], or iteratively using a refined dictionary [49]. An alternative is to consider the model (2.4) directly as in [46, 124]. However, such approaches cannot be generalized to scenarios in which the analog signals contain several bands, each with non-negligible bandwidth.

2.4.2 Multiband Signals

A more realistic model for a structured analog signal is a *multiband* model, in which $x(t)$ has a CTFT supported on a union of several narrow bands

$$\mathbb{F} = \bigcup_{i=0}^{J-1} [F_i - B_{\text{band}_i}/2, F_i + B_{\text{band}_i}/2],$$

i.e.,

$$x(t) = \int_{\mathbb{F}} X(F) e^{j2\pi Ft} dF.$$

Here $X(F)$ denotes CTFT of $x(t)$. The band centers are given by the frequencies $\{F_i\}_{i\in[J]}$ and the band widths are denoted by $\{B_{\text{band}_i}\}_{i\in[J]}$, where $[J]$ denotes the set $\{0, 1, \ldots, J-1\}$.

Again we let $\boldsymbol{x}$, as defined in (2.3), denote the length-N vector obtained by uniformly sampling $x(t)$ over the time interval $[0, NT_s)$ with sampling period T_s. We assume T_s is chosen to satisfy the minimum Nyquist sampling rate, which means

$$T_s \leq \frac{1}{B_{\text{nyq}}} := \frac{1}{2\max_{i\in[J]} \{|F_i \pm B_{\text{band}_i}/2|\}}.$$

Under these assumptions, the sampled multiband signal $\boldsymbol{x}$ can be expressed as an integral of sampled pure tones (i.e., discrete-time sinusoids)

$$\boldsymbol{x}[n] = \int_{\mathbb{W}} \widetilde{x}(f) e^{j2\pi fn}\, df,\ n = 0, 1, \ldots, N-1, \tag{2.5}$$

where the digital frequency f is integrated over the union of intervals

$$\mathbb{W} := [f_0 - W_0, f_0 + W_0] \cup [f_1 - W_1, f_1 + W_1] \cup \cdots \cup [f_{J-1} - W_{J-1}, f_{J-1} + W_{J-1}] \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right] \quad (2.6)$$

with $f_i = T_s F_i$ and $W_i = T_s B_{\text{band}_i}/2$ for all $i \in [J]$. The weighting function $\widetilde{x}(f)$ appearing in (2.5) equals the scaled CTFT of $x(t)$,

$$\widetilde{x}(f) = \frac{1}{T_s} X(F)|_{F=\frac{f}{T_s}}, \ |f| \leq \frac{1}{2},$$

and corresponds to the DTFT of the infinite sample sequence $\{\ldots, x(-T_s), x(0), x(T_s), \ldots\}$. (However, we stress that our interest is on the finite-length sample vector $\boldsymbol{x}$ and not on this infinite sample sequence.) Such multiband signal models arise in problems such as radar signal processing with non-point targets [2] and mitigation of narrowband interference [30, 33].

2.5 Time, Index, and Multiband-Limiting Operators and the Prolate Matrix

Let $\mathcal{B}_{\mathbb{W}} : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ denote the multiband-limiting operator that bandlimits the DTFT of a discrete-time signal to the frequency range $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$, i.e., for $y \in \ell_2(\mathbb{Z})$, we have that[1]

$$\mathcal{B}_{\mathbb{W}}(y)[m] := \int_{\mathbb{W}} e^{j2\pi f m} df \star y[m] = \sum_{n=-\infty}^{\infty} \left(y[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), \quad (2.7)$$

where $\star$ stands for convolution. In addition, let $\mathcal{T}_N : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ denote the operator that zeros out all entries outside the index range $\{0, 1, \ldots, N-1\}$. That is

$$\mathcal{T}_N(y)[m] := \begin{cases} y[m], & m \in [N], \\ 0, & \text{otherwise.} \end{cases}$$

Next, define the index-limiting operator $\mathcal{I}_N : \ell_2(\mathbb{Z}) \to \mathbb{C}^N$ as

$$\mathcal{I}_N(y)[m] := y[m], \ m \in [N].$$

The adjoint operator $\mathcal{I}_N^* : \mathbb{C}^N \to \ell_2(\mathbb{Z})$ (anti-index-limiting operator) is given by

[1]For convenience, we use $\mathcal{B}_W$ instead of $\mathcal{B}_{[-W,W]}$ when $\mathbb{W}$ reduces to $[-W, W]$. This is also the reason for many other notations involving subscript W.

$$\mathcal{I}_N^*(\boldsymbol{y})[m] := \begin{cases} \boldsymbol{y}[m], & m \in [N], \\ 0, & \text{otherwise.} \end{cases}$$

We can observe that $\mathcal{T}_N = \mathcal{I}_N^* \mathcal{I}_N$. By an abuse of notation, we assume that $\mathcal{I}_N$ also works on finite-dimensional vectors, i.e., for any $\boldsymbol{y} \in \mathbb{C}^M$ with $M \geq N$, $\mathcal{I}_N(\boldsymbol{y}) \in \mathbb{C}^N$ with elements given by

$$\mathcal{I}_N(\boldsymbol{y})[m] := \boldsymbol{y}[m], \ m \in [N].$$

The input of $\mathcal{I}_N$ should be clear from the context.

Now the time- and multiband-limiting operator $\mathcal{B}_{\mathbb{W}}\mathcal{T}_N : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ is defined by

$$\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y))[m] := \sum_{n=0}^{N-1} \left(y[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), \ m \in \mathbb{Z}. \tag{2.8}$$

Further composing the time- and multiband-limiting operators, we obtain the linear operator $\mathcal{T}_N \mathcal{B}_{\mathbb{W}} \mathcal{T}_N : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ as

$$\mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y)))[m] = \begin{cases} \sum_{n=0}^{N-1} \left(y[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), & m \in [N], \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Similarly, combining the index- and multiband-limiting operators, we obtain the linear operator $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^* : \mathbb{C}^N \to \mathbb{C}^N$ as

$$\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y})))[m] = \sum_{n=0}^{N-1} \left(\boldsymbol{y}[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), \ m \in [N]. \tag{2.10}$$

Suppose $y' \in \ell_2(\mathbb{Z})$ is an eigenfunction of $\mathcal{T}_N \mathcal{B}_{\mathbb{W}} \mathcal{T}_N$ with corresponding eigenvalue λ': $\mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y'))) = \lambda' y'$. We can verify that $\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\mathcal{I}_N(y')))) = \lambda' \mathcal{I}_N(y')$. On the other hand, if $\boldsymbol{y}''$ and λ'' satisfy $\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y}''))) = \lambda'' \boldsymbol{y}''$, then we can conclude that $\mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(\mathcal{I}_N^*(\boldsymbol{y}'')))) = \lambda'' \mathcal{I}_N^*(\boldsymbol{y}'')$. Therefore $\mathcal{T}_N \mathcal{B}_{\mathbb{W}} \mathcal{T}_N$ and $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ have the same eigenvalues, and the eigenvectors of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ can be obtained by index-limiting the eigenvectors of $\mathcal{T}_N \mathcal{B}_{\mathbb{W}} \mathcal{T}_N$.

It is easy to show that $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ is equivalent to the *prolate matrix* $\boldsymbol{B}_{N,\mathbb{W}}$, which has entries

$$\boldsymbol{B}_{N,\mathbb{W}}[m,n] := \int_{\mathbb{W}} e^{j2\pi f(m-n)} df = \sum_{i=0}^{J-1} e^{j2\pi f_i(m-n)} \frac{\sin\left(2\pi W_i(m-n)\right)}{\pi(m-n)} \tag{2.11}$$

for all $m, n \in [N]$.

2.6 The Slepian Basis and the Fourier Basis

In this section, we provide a formal definition of the Slepian basis and briefly describe some of the key results from Slepian's 1978 paper on DPSS's [118]. Given any $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$, the DPSS's are a collection of N discrete-time sequences that are strictly bandlimited to the digital frequency range $|f| \leq W$ yet highly concentrated in time to the index range $n = 0, 1, \ldots, N-1$. The DPSS's are defined to be the eigenvectors of a two-step procedure in which one first time-limits the sequence and then bandlimits the sequence. Recall that $\mathcal{B}_W$ denotes an operator that takes a discrete-time signal, bandlimits its DTFT to the frequency range $|f| \leq W$, and returns the corresponding signal in the time domain. Additionally, $\mathcal{T}_N$ is an operator that takes an infinite-length discrete-time signal and zeros out all entries outside the index range $\{0, 1, \ldots, N-1\}$ (but still returns an infinite-length signal). With these definitions, the DPSS's are defined in [118] as follows.

Definition 2.1. *Given any $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$, the DPSS's are a collection of N real-valued discrete-time sequences $s_{N,W}^{(0)}, s_{N,W}^{(1)}, \ldots, s_{N,W}^{(N-1)}$ that, along with the corresponding scalar eigenvalues $1 > \lambda_{N,W}^{(0)} > \lambda_{N,W}^{(1)} > \cdots > \lambda_{N,W}^{(N-1)} > 0$, satisfy*

$$\mathcal{B}_W(\mathcal{T}_N(s_{N,W}^{(\ell)})) = \lambda_{N,W}^{(\ell)} s_{N,W}^{(\ell)} \tag{2.12}$$

for all $\ell \in \{0, 1, \ldots, N-1\}$. The DPSS's are normalized so that

$$\|\mathcal{T}_N(s_{N,W}^{(\ell)})\|_2 = 1 \tag{2.13}$$

for all $\ell \in \{0, 1, \ldots, N-1\}$.

One of the central contributions of [118] was to examine the behavior of the eigenvalues $\lambda_{N,W}^{(0)}, \ldots, \lambda_{N,W}^{(N-1)}$. In particular, [118] shows that the first $2NW$ eigenvalues tend to cluster

extremely close to 1, while the remaining eigenvalues tend to cluster similarly close to 0. This is made more precise in the following lemma from [118].

Lemma 2.1. *(Clustering of eigenvalues [118, 32]) Suppose that $W \in (0, \frac{1}{2})$ is fixed.*

1. *Fix $\epsilon \in (0, 1)$. Then there exist constants $C_1(W, \epsilon), C_2(W, \epsilon)$ (which may depend on W, ϵ) and an integer $N_0(W, \epsilon)$ (which may also depend on W, ϵ) such that*

$$1 - \lambda_{N,W}^{(\ell)} \leq C_1(W, \epsilon) e^{-C_2(W,\epsilon)N}, \ \forall \ \ell \leq \lfloor 2NW(1-\epsilon) \rfloor$$

   for all $N \geq N_0(W, \epsilon)$.

2. *Fix $\epsilon \in (0, \frac{1}{2W} - 1)$. Then there exist constants $C_3(W, \epsilon), C_4(W, \epsilon)$ (which may depend on W, ϵ) and an integer $N_1(W, \epsilon)$ (which may also depend on W, ϵ) such that*

$$\lambda_{N,W}^{(\ell)} \leq C_3(W, \epsilon) e^{-C_4(W,\epsilon)N}, \ \forall \ \ell \geq \lceil 2NW(1+\epsilon) \rceil$$

   for all $N \geq N_1(W, \epsilon)$.

This tells us that the range of the operator $\mathcal{B}_W \mathcal{T}_N$ has an effective dimension of $\approx 2NW$. Moreover, with only a few exceptions near the "transition region" at $\ell \approx 2NW$, we can reasonably approximate the eigenvalues $\lambda_{N,W}^{(\ell)}$ to be either 1 or 0. This will play a central role throughout our analysis.

Finally, we also note that while each DPSS actually has infinite support in time, several very useful properties hold for the collection of signals one obtains by time-limiting the DPSS's to the index range $n = 0, 1, \ldots, N-1$. First, it can be shown that [118]

$$\|\mathcal{B}_W(\mathcal{T}_N(s_{N,W}^{(\ell)}))\|_2 = \sqrt{\lambda_{N,W}^{(\ell)}}. \tag{2.14}$$

Comparing (2.13) with (2.14), we see that for values of ℓ where $\lambda_{N,W}^{(\ell)} \approx 1$, nearly all of the energy in $\mathcal{T}_N(s_{N,W}^{(\ell)})$ is contained in the frequencies $|f| \leq W$. While by construction the DTFT of any DPSS is perfectly bandlimited, the DTFT of the corresponding time-limited DPSS will only be concentrated in the bandwidth of interest for the first $\approx 2NW$ DPSS's. As a result, we will frequently be primarily interested in roughly the first $2NW$ DPSS's.

Second, the time-limited DPSS's are orthogonal [118] so that for any $\ell, \ell' \in \{0, 1, \ldots, N-1\}$ with $\ell \neq \ell'$,

$$\langle \mathcal{T}_N(s_{N,W}^{(\ell)}), \mathcal{T}_N(s_{N,W}^{(\ell')}) \rangle = 0. \tag{2.15}$$

Finally, like the DPSS's, the time-limited DPSS's have a special eigenvalue relationship with the time-limiting and bandlimiting operators. In particular, if we apply the operator $\mathcal{T}_N$ to both sides of (2.12), we see that the sequences $\mathcal{T}_N(s_{N,W}^{(\ell)})$ are actually eigenfunctions of the two-step procedure in which one first bandlimits a sequence and then time-limits the sequence.

These properties, together with the fact that our focus is primarily on providing computational tools for finite-length vectors, motivate our definition of the Slepian basis to be the restriction of the (time-limited) DPSS's to the index range $n = 0, 1, \ldots, N-1$ (discarding the zeros outside this range).

Definition 2.2. *Given any $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$, the Slepian basis is given by the vectors $\boldsymbol{s}_{N,W}^{(0)}, \boldsymbol{s}_{N,W}^{(1)}, \ldots, \boldsymbol{s}_{N,W}^{(N-1)} \in \mathbb{R}^N$ which are defined by restricting the time-limited DPSS's to the index range $n = 0, 1, \ldots, N-1$:*

$$\boldsymbol{s}_{N,W}^{(\ell)}[n] := \mathcal{T}_N(s_{N,W}^{(\ell)})[n] = s_{N,W}^{(\ell)}[n]$$

for all $\ell, n \in \{0, 1, \ldots, N-1\}$. For simplicity, we will often use the notation $\boldsymbol{S}_{N,W}$ to denote the $N \times N$ matrix given by

$$\boldsymbol{S}_{N,W} = \begin{bmatrix} \boldsymbol{s}_{N,W}^{(0)} & \cdots & \boldsymbol{s}_{N,W}^{(N-1)} \end{bmatrix}.$$

Observe that combining (2.13) and (2.15), it follows that $\boldsymbol{S}_{N,W}$ does indeed form an orthonormal basis for $\mathbb{C}^N$ (or for $\mathbb{R}^N$). However, following from our discussion above, the partial Slepian basis constructed using just the first $\approx 2NW$ basis elements will play a special role and can be shown to be remarkably effective for capturing the energy in a length-N window of samples of a bandlimited signal (see [32] for further discussion.) In such

situations, we will also use the notation $\boldsymbol{S}_K$ to denote the first K columns of $\boldsymbol{S}_{N,W}$ (where N and W are clear from the context and typically $K \approx 2NW$).

In our discussion above we derived the Slepian basis by following the same approach as in [118] and considering the time-limitations of the eigenfunctions of the operator given by $\mathcal{B}_W \mathcal{T}_N$. As illustrated in (2.11), an alternative way to derive $\boldsymbol{S}_{N,W}$ is to consider the eigenvectors of the $N \times N$ prolate matrix $\boldsymbol{B}_{N,W}$ [132], which is the matrix with entries given by

$$\boldsymbol{B}_{N,W}[m,n] := \frac{\sin 2\pi W(m-n)}{\pi(m-n)} \tag{2.16}$$

for all $m, n \in \{0, 1, \ldots, N-1\}$. Indeed, $\boldsymbol{B}_{N,W}$ can be understood as the finite truncation of the infinite matrix representation of $\mathcal{B}_W \mathcal{T}_N$. Thus, $\boldsymbol{S}_{N,W}$ contains the eigenvectors of $\boldsymbol{B}_{N,W}$ and we can write $\boldsymbol{B}_{N,W}$ as

$$\boldsymbol{B}_{N,W} = \boldsymbol{S}_{N,W} \boldsymbol{\Lambda}_{N,W} \boldsymbol{S}_{N,W}^*$$

where $\boldsymbol{\Lambda}_{N,W}$ is an $N \times N$ diagonal matrix with the eigenvalues $\lambda_{N,W}^{(0)}, \ldots, \lambda_{N,W}^{(N-1)}$, along the main diagonal (sorted in descending order).

The primary goal of Chapter 3 is to develop fast algorithms for working with $\boldsymbol{S}_{N,W}$ (or $\boldsymbol{B}_{N,W}$, which also arises in many practical applications, as detailed in Section 3.3 below). Towards this end, we will begin by examining the relationship between $\boldsymbol{B}_{N,W}$ and the matrix obtained by projecting onto the lowest $2NW$ Fourier coefficients. To be more precise, we define W' such that $2NW'$ is the nearest odd integer to $2NW$, and we let $\boldsymbol{F}_{N,W}$ denote the partial Fourier matrix with the lowest $2NW'$ frequency DFT vectors of length N, i.e.,

$$\boldsymbol{F}_{N,W} = \frac{1}{\sqrt{N}} \begin{bmatrix} \boldsymbol{e}_{-(2NW'-1)/2N} & \cdots & \boldsymbol{e}_{(2NW'-1)/2N} \end{bmatrix}. \tag{2.17}$$

Note that the projection onto the span of $\boldsymbol{F}_{N,W}$ is given by the matrix $\boldsymbol{F}_{N,W} \boldsymbol{F}_{N,W}^*$, which has entries given by

$$[\boldsymbol{F}_{N,W} \boldsymbol{F}_{N,W}^*][m,n] = \frac{1}{N} \sum_{k=-NW'+\frac{1}{2}}^{NW'-\frac{1}{2}} e^{j2\pi(m-n)k/N} = \frac{\sin(2\pi W'(m-n))}{N \sin(\pi \frac{m-n}{N})} \tag{2.18}$$

for $m, n = 0, \ldots, N-1$.

2.7 Representations of Sampled Sinusoids and Oversampled Bandlimited Signals

Note that for any orthonormal matrix $\boldsymbol{Q} \in \mathbb{C}^{N \times K}$,

$$\begin{aligned} \int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2 \, df &= \int_{-W}^{W} \operatorname{trace}\left(\boldsymbol{e}_f \boldsymbol{e}_f^* - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f \boldsymbol{e}_f^*\right) df \\ &= \operatorname{trace}\left(\boldsymbol{B}_{N,W} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{B}_{N,W}\right), \end{aligned} \tag{2.19}$$

where $\boldsymbol{e}_f$ is a length-N sinusoid defined in (2.1). For any value of K, the quantity in (2.19) is minimized by the choice of $\boldsymbol{Q} = \boldsymbol{S}_K$. This implies that $\boldsymbol{S}_K$ is the best basis of K columns to represent (in a MSE sense) the collection of sampled sinusoids $\{\boldsymbol{e}_f\}_{f \in [-W,W]}$. Formally,

$$\int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{e}_f\|_2^2 \, df = \sum_{\ell=K}^{N-1} \lambda_{N,W}^{(\ell)}, \tag{2.20}$$

whereas for each $f \in [-W, W]$, $\|\boldsymbol{e}_f\|_2^2 = N$. It follows from Lemma 2.1 that $\boldsymbol{S}_K$ provides very accurate approximations (in a MSE sense) for all sampled sinusoids $\{\boldsymbol{e}_f\}_{f \in [-W,W]}$ if one chooses K slightly larger than $2NW$.

We note that any representation guarantee for sampled sinusoids $\{\boldsymbol{e}_f\}_{f \in [-W,W]}$ can also be used for finite-length sample vectors arising from sampling random bandlimited baseband signals. Suppose x is a continuous-time, zero-mean, wide sense stationary random process with power spectrum

$$P_x(F) = \begin{cases} \frac{1}{B_{\text{band}}}, & F \in [-\frac{B_{\text{band}}}{2}, \frac{B_{\text{band}}}{2}], \\ 0, & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{x} = [x(0) \; x(T_s) \; \cdots \; x((N-1)T_s)]^T \in \mathbb{C}^N$ denote a finite vector of samples acquired from $x(t)$ with a sampling interval of $T_s \leq 1/B_{\text{band}}$. Let $f_c = F_c T_s$ and $W = \frac{B_{\text{band}} T_s}{2}$. We have [32]

$$\mathbb{E}\left[\|\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{x}\|_2^2\right] = \frac{1}{2W} \int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2 \, df. \tag{2.21}$$

2.8 Szegő's Theorem

Toeplitz matrices are of considerable interest in statistical signal processing and information theory [56, 54, 104, 91, 73]. An $N \times N$ Toeplitz matrix $\boldsymbol{H}_N$ has the form

$$\boldsymbol{H}_N = \begin{bmatrix} h[0] & h[-1] & h[-2] & \dots & h[-(N-1)] \\ h[1] & h[0] & h[-1] & & \\ h[2] & h[1] & h[0] & & \vdots \\ \vdots & & & \ddots & \\ h[N-1] & & \cdots & & h[0] \end{bmatrix} \tag{2.22}$$

or $\boldsymbol{H}_N[m,n] = h[m-n]; m,n \in [N] := \{0,1,\dots,N-1\}$. The covariance matrix of a discrete-time wide-sense stationary (WSS) random process is an example of such a matrix. Also the prolate matrices (defined in (2.11)) are Toeplitz.

Throughout this thesis, we consider $\boldsymbol{H}_N$ that is Hermitian, i.e., $\boldsymbol{H}_N^* = \boldsymbol{H}_N$, and we suppose that the eigenvalues of $\boldsymbol{H}_N$ are denoted and arranged as $\lambda_0(\boldsymbol{H}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{H}_N)$. Here the Hermitian transpose of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^H$.

Szegő's theorem [56] describes the collective asymptotic behavior (as $N \to \infty$) of the eigenvalues of a sequence of Hermitian Toeplitz matrices $\{\boldsymbol{H}_N\}$ by relating to its DTFT $\widetilde{h}(f) \in L^2([0,1])$:

$$\begin{aligned} h[k] &= \int_0^1 \widetilde{h}(f) e^{-j2\pi kf} df, \quad k \in \mathbb{Z}, \\ \widetilde{h}(f) &= \sum_{k=-\infty}^{\infty} h[k] e^{j2\pi kf}, \quad f \in [0,1]. \end{aligned}$$

Usually $\widetilde{h}(f)$ is referred to as the symbol or generating function for the $N \times N$ Toeplitz matrices $\{\boldsymbol{H}_N\}$.

Suppose $\widetilde{h} \in L^\infty([0,1])$. Szegő's theorem [56] states that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widetilde{h}(f)) df, \tag{2.23}$$

where ϑ is any function continuous on the range of $\widetilde{h}$. As one example, choosing $\vartheta(x) = x$ yields

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \lambda_l(\boldsymbol{H}_N) = \int_0^1 \widetilde{h}(f) df.$$

In words, this says that as $N \to \infty$, the average eigenvalue of $\boldsymbol{H}_N$ converges to the average value of the symbol $\widetilde{h}(f)$ that generates $\boldsymbol{H}_N$. As a second example, suppose $\widetilde{h}(f) > 0$ (and thus $\lambda_l(\boldsymbol{H}_N) > 0$ for all $l \in [N]$ and $N \in \mathbb{N}$) and let ϑ be the log function. Then Szegő's theorem indicates that

$$\lim_{N\to\infty} \frac{1}{N} \log\left(\det\left(\boldsymbol{H}_N\right)\right) = \int_0^1 \log\left(\widetilde{h}(f)\right) df.$$

This relates the determinant of the Toeplitz matrix to its symbol.

Szegő's theorem has been widely used in the areas of signal processing, communications, and information theory. A paper and review by Gray [54, 55] serve as a remarkable elementary introduction in the engineering literature and offer a simplified proof of Szegő's original theorem. The result has also been extended in several ways. For example, the Avram-Parter theorem [6, 103], a generalization of Szegő's theorem, relates the collective asymptotic behavior of the singular values of a general (non-Hermitian) Toeplitz matrix to the absolute value of its symbol, i.e., $|\widetilde{h}(f)|$. Tyrtyshnikov [130] proved that Szegő's theorem holds if $\widetilde{h}(f) \in \mathbb{R}$ and $\widetilde{h}(f) \in L^2([0,1])$, and Zamarashkin and Tyrtyshnikov [139] further extended Szegő's theorem to the case when $\widetilde{h}(f) \in \mathbb{R}$ and $\widetilde{h}(f) \in L^1([0,1])$. Sakrison [108] extended Szegő's theorem to high dimensions. Gazzah et al. [52] and Gutiérrez-Gutiérrez and Crespo [59] extended Gray's results on Toeplitz and circulant matrices to block Toeplitz and block circulant matrices and derived Szegő's theorem for block Toeplitz matrices.

2.9 Subspace Angle

To quantify the "distance" between two subspaces with possible different dimensions, we establish the following definition of angle between subspaces to compare subspaces of possibly

different dimensions.

Definition 2.3. *Let $\mathcal{S}_{\boldsymbol{A}}$ and $\mathcal{S}_{\boldsymbol{B}}$ be the subspaces formed by the columns of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. The subspace angle $\Theta_{\boldsymbol{A},\boldsymbol{B}}$ between $\mathcal{S}_{\boldsymbol{A}}$ and $\mathcal{S}_{\boldsymbol{B}}$ is given by*

$$\cos(\Theta_{\boldsymbol{A},\boldsymbol{B}}) := \inf_{\boldsymbol{a}\in\mathcal{S}_{\boldsymbol{A}},\|\boldsymbol{a}\|_2=1} \|\boldsymbol{P}_{\boldsymbol{B}}\boldsymbol{a}\|_2$$

if $\dim(\mathcal{S}_{\boldsymbol{B}}) \geq \dim(\mathcal{S}_{\boldsymbol{A}})$*, or*

$$\cos(\Theta_{\boldsymbol{A},\boldsymbol{B}}) := \inf_{\boldsymbol{b}\in\mathcal{S}_{\boldsymbol{B}},\|\boldsymbol{b}\|_2=1} \|\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{b}\|_2$$

if $\dim(\mathcal{S}_{\boldsymbol{B}}) < \dim(\mathcal{S}_{\boldsymbol{A}})$*. Here $\boldsymbol{P}_{\boldsymbol{B}}$ (or $\boldsymbol{P}_{\boldsymbol{A}}$) denotes the orthogonal projection onto the column space of $\boldsymbol{B}$ (or $\boldsymbol{A}$).*

We remark that when the subspaces $\mathcal{S}_{\boldsymbol{A}}$ and $\mathcal{S}_{\boldsymbol{B}}$ have the same dimension, our definition of subspace angle coincides to the *subspace gap* [43], defined as $\sin(\Theta_{\boldsymbol{A},\boldsymbol{B}})$. Smaller $\Theta_{\boldsymbol{A},\boldsymbol{B}}$ indicates a smaller gap between $\mathcal{S}_{\boldsymbol{A}}$ and $\mathcal{S}_{\boldsymbol{B}}$. We also connect our definition of subspace angle to *principal angles* between two subspaces defined as follows.

Definition 2.4. *[10] Suppose $\boldsymbol{A} \in \mathbb{R}^{N\times p}$ and $\boldsymbol{B} \in \mathbb{R}^{N\times q}$ are orthonormal bases for the subspaces $\mathcal{S}_{\boldsymbol{A}} \subset \mathbb{R}^{N\times N}$ and $\mathcal{S}_{\boldsymbol{B}}$, respectively. Suppose $p \geq q$. Then the principal angles between $\mathcal{S}_{\boldsymbol{A}}$ and $\mathcal{S}_{\boldsymbol{B}}$, $\phi_1(\boldsymbol{A},\boldsymbol{B}) \leq \phi_2(\boldsymbol{A},\boldsymbol{B}) \leq \cdots \leq \phi_q(\boldsymbol{A},\boldsymbol{B})$, are defined as*

$$\cos\left(\phi_i(\boldsymbol{A},\boldsymbol{B})\right) = \sigma_i(\boldsymbol{A}^*\boldsymbol{B})$$

for all $i \in \{1,2,\ldots,q\}$, where $\sigma_i(\cdot)$ denotes the i-th largest singular value.

We note that the subspace angle $\Theta_{\boldsymbol{A},\boldsymbol{B}}$ is equivalent to the largest principal angle $\phi_q(\boldsymbol{A},\boldsymbol{B})$. To see this, we rewrite the smallest singular value:

$$\cos\left(\phi_q(\boldsymbol{A},\boldsymbol{B})\right) = \sigma_q(\boldsymbol{A}^*\boldsymbol{B}) = \inf_{\|\boldsymbol{\alpha}\|_2=1} \|\boldsymbol{A}^*\boldsymbol{B}\boldsymbol{\alpha}\|_2 = \inf_{\boldsymbol{b}\in\mathcal{S}_{\boldsymbol{B}},\|\boldsymbol{b}\|_2=1,} \|\boldsymbol{A}^*\boldsymbol{b}\|_2 = \inf_{\boldsymbol{b}\in\mathcal{S}_{\boldsymbol{B}},\|\boldsymbol{b}\|_2=1,} \|\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{b}\|_2\,,$$

where the last inequality follows because by assumption $\boldsymbol{A}$ is an orthonormal basis for $\mathcal{S}_{\boldsymbol{A}}$. Thus, our definition of subspace angle captures the largest possible principal angle between two subspaces.

2.10 Harmonic Analysis on Locally Compact Abelian Groups

We now provide some basic terminology in group theory and introduce the generalized Fourier transforms for functions defined on groups.

2.10.1 Groups and Dual Groups

To begin, we first list some necessary definitions related to groups.

Definition 2.5 (Definition 7.1 [25]). *A (closed) binary operation, $\circ$, is a law of composition that produces an element of a set from two elements of the same set. More precisely, let $\mathbb{G}$ be a set and $g_1, g_2 \in \mathbb{G}$ be arbitrary elements. Then $(g_1, g_2) \to g_1 \circ g_2 \in \mathbb{G}$.*

Definition 2.6 (Definition 7.2 [25]). *A* **group** *is a set $\mathbb{G}$ together with a (closed) binary operation $\circ$ such that the following properties hold:*

- *Associative property: $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$ holds for any $g_1, g_2, g_3 \in \mathbb{G}$.*
- *There exists an identity element $e \in \mathbb{G}$ such that $e \circ g = g \circ e = g$ holds for all $g \in \mathbb{G}$.*
- *For any $g \in \mathbb{G}$, there is an element $g^{-1} \in \mathbb{G}$ such that $g^{-1} \circ g = g \circ g^{-1} = e$.*

With this definition, it is common to denote a group just by $\mathbb{G}$ without mentioning the binary operation $\circ$ when it is clear from the context.

Let $\mathbb{G}$ denote a *locally compact abelian group*.[2] A locally compact abelian group can be either discrete or continuous, and either compact or non-compact. A *character* $\chi_\xi : \mathbb{G} \to \mathbb{T}$ of $\mathbb{G}$ is a continuous group homomorphism from $\mathbb{G}$ with values in the circle group $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ satisfying

[2]To simplify many technical details, we only consider locally compact abelian groups. A locally compact group is a topological group for which the underlying topology is locally compact and Hausdorff (which is a topological space in which distinct points have disjoint neighborhoods). An abelian group, also called a commutative group, is a group in which the result of applying the group operation to two group elements does not depend on the order. When $\mathbb{G}$ is locally compact but neither compact nor abelian, many of our results still hold but become more complex. For example, even choosing a suitable measure on $\widehat{\mathbb{G}}$ for a general $\mathbb{G}$ is a difficult problem. Only under appropriate conditions can one find an appropriate measure on $\widehat{\mathbb{G}}$ such that the inversion formula holds.

$$
\begin{aligned}
&|\chi_\xi(g)| = 1,\\
&\chi_\xi^*(g) = \chi_\xi(g^{-1}),\\
&\chi_\xi(h \circ g) = \chi_\xi(h)\chi_\xi(g).
\end{aligned}
$$

for any $g, h \in \mathbb{G}$. Here $\chi_\xi^*(g)$ is the complex conjugate of $\chi_\xi(g)$. The set of all characters on $\mathbb{G}$ introduces a locally compact abelian group, called the *dual group* of $\mathbb{G}$ and denoted by $\widehat{\mathbb{G}}$ if we pair $(g, \xi) \to \chi_\xi(g)$ for all $\xi \in \widehat{\mathbb{G}}$ and $g \in \mathbb{G}$. In most references the character is denoted simply by χ rather than by χ_ξ. However, we use here the notation χ_ξ in order to emphasize that the character can be viewed as a function of two elements $g \in \mathbb{G}$ and $\xi \in \widehat{\mathbb{G}}$, and for any $\xi \in \widehat{\mathbb{G}}$, χ_ξ is a function of g. In this sense, $\chi_\xi(g)$ can be regarded as the value of the character χ_ξ evaluated at the group element g. Table 2.1 lists several examples of groups $\mathbb{G}$, the corresponding binary operation $\circ$ and the corresponding dual groups $\widehat{\mathbb{G}}$ that have relevance in signal processing and information theory. Here $\mathrm{mod}(a, b) = \frac{a}{b} - \lfloor \frac{a}{b} \rfloor$, where $\lfloor c \rfloor$ is the largest integer that is not greater than c.

Table 2.1: Examples of groups $\mathbb{G}$, along with their dual groups $\mathbb{G}$ and Fourier transforms.

group $\mathbb{G}$	dual group $\widehat{\mathbb{G}}$	g	binary operation $\circ$	ξ	$\chi_\xi(g)$
$\mathbb{R}$	$\mathbb{R}$	t	$t_1 + t_2$	F	$e^{j2\pi Ft}$
$\mathbb{R}^n$	$\mathbb{R}^n$	$\boldsymbol{t}$	$\boldsymbol{t}_1 + \boldsymbol{t}_2$	$\boldsymbol{F}$	$e^{j2\pi\langle \boldsymbol{F}, \boldsymbol{t}\rangle}$
unit circle $[0, 1)$	$\mathbb{Z}$	t	$\mathrm{mod}(t_1 + t_2, 1)$	k	$e^{j2\pi tk}$
$\mathbb{Z}$	unit circle	n	$n_1 + n_2$	f	$e^{j2\pi fn}$
$\mathbb{Z}_N = N$ roots of unity	$\mathbb{Z}_N = N$ roots of unity	n	$\mathrm{mod}(n_1 + n_2, N)$	k	$e^{j2\pi \frac{nk}{N}}$

2.10.2 Fourier Transforms

The characters $\{\chi_\xi\}_{\xi \in \widehat{\mathbb{G}}}$ play an important role in defining the Fourier transform for functions in $L_2(\mathbb{G})$. In particular, the Pontryagin duality theorem [106], named after Lev Semennovich Pontryagin who laid down the foundation for the theory of locally compact abelian groups, generalizes the conventional CTFT on $L_2(\mathbb{R})$ and CT Fourier series for

periodic functions to functions defined on locally compact abelian groups.

Theorem 2.1 (Pontryagin duality theorem [106]). *Let $\mathbb{G}$ be a locally compact abelian group and μ be a Haar measure on $\mathbb{G}$. Let $x(g) \in L_2(\mathbb{G})$. Then the Fourier transform $\widehat{x}(\xi) \in L_2(\widehat{\mathbb{G}})$ is defined by*

$$\widehat{x}(\xi) = \int_{\mathbb{G}} x(g)\chi_\xi^*(g)\,\mathrm{d}\,\mu(g).$$

For each Haar measure μ on $\mathbb{G}$ there is a unique Haar measure ν on $\widehat{\mathbb{G}}$ such that the following inverse Fourier transform holds

$$x(g) = \int_{\widehat{\mathbb{G}}} \widehat{x}(\xi)\chi_\xi(g)\,\mathrm{d}\,\nu(\xi).$$

The Fourier transform satisfies Parseval's theorem:

$$\int_{\mathbb{G}} |x(g)|^2\,\mathrm{d}\,\mu(g) = \int_{\widehat{\mathbb{G}}} |\widehat{x}(\xi)|^2\,\mathrm{d}\,\nu(\xi).$$

Only Haar measure and integral on $\mathbb{G}$ are considered throughout this paper. We note that the unique Haar measure ν on $\widehat{\mathbb{G}}$ depends on the choice of Haar measure μ on $\mathbb{G}$. We illustrate this point with the conventional DFT as an example where $g = n \in \mathbb{G} = \mathbb{Z}_N$, $\xi = k \in \widehat{\mathbb{G}} = \mathbb{Z}_N$, and $\chi_\xi(g) = e^{j2\pi\frac{nk}{N}}$. If we choose the counting measure (where each element of $\mathbb{G}$ receives a value of 1) on $\mathbb{G}$, then we must use the normalized counting measure (where each element of $\widehat{\mathbb{G}}$ receives a value of $\frac{1}{N}$) on $\widehat{\mathbb{G}}$. The DFT and inverse DFT become

$$\widehat{\boldsymbol{x}}[k] = \sum_{n=0}^{N-1} \boldsymbol{x}[n]e^{-j2\pi\frac{nk}{N}};\ \boldsymbol{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \widehat{\boldsymbol{x}}[k]e^{j2\pi\frac{nk}{N}}.$$

One can also choose the semi-normalized counting measure (where each element receives a value of $\frac{1}{\sqrt{N}}$) on both groups $\mathbb{G}$ and $\widehat{\mathbb{G}}$. This gives the normalized DFT and inverse DFT:

$$\widehat{\boldsymbol{x}}[k] = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} \boldsymbol{x}[n]e^{-j2\pi\frac{nk}{N}};\ \boldsymbol{x}[n] = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} \widehat{\boldsymbol{x}}[k]e^{j2\pi\frac{nk}{N}}.$$

For convenience, we rewrite the Fourier transform and inverse Fourier transform as follows when the Haar measures are clear from the context:

$$\widehat{x}(\xi) = \int_{\mathbb{G}} x(g)\chi_\xi^*(g)\,\mathrm{d}\,g;\ x(g) = \int_{\widehat{\mathbb{G}}} \widehat{x}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi.$$

We also use $\mathcal{F} : L_2(\mathbb{G}) \to L_2(\widehat{\mathbb{G}})$ and $\mathcal{F}^{-1} : L_2(\widehat{\mathbb{G}}) \to L_2(\mathbb{G})$ to denote the operators corresponding to the Fourier transform and inverse Fourier transform, respectively.

For each group $\mathbb{G}$ and dual group $\widehat{\mathbb{G}}$ listed in Table 2.1, the table also includes the corresponding Fourier transform.

2.10.3 Convolutions

For any $x(g), y(g) \in L_2(\mathbb{G})$, we define the convolution between $x(g)$ and $y(g)$ by

$$(x \star y)(g) := \int_{\mathbb{G}} y(h)x(h^{-1} \circ g)\,\mathrm{d}\,h. \tag{2.24}$$

Similar to what holds in the standard CT and DT signal processing contexts, it is not difficult to show that the Fourier transform on $\mathbb{G}$ also takes convolution to multiplication. That is, for any $x, y \in L_2(\mathbb{G})$,

$$\begin{aligned}
\mathcal{F}(x \star y)(\xi) &= \int_{\mathbb{G}} \int_{\mathbb{G}} y(h)x(h^{-1} \circ g)\,\mathrm{d}\,h\ \chi_\xi^*(g)\,\mathrm{d}\,g \\
&= \int_{\mathbb{G}} \int_{\mathbb{G}} x(h^{-1} \circ g)\chi_\xi^*(h^{-1} \circ g)\,\mathrm{d}\,g\ \chi_\xi^*(h)y(h)\,\mathrm{d}\,h \\
&= (\mathcal{F}x)(\xi)(\mathcal{F}y)(\xi)
\end{aligned}$$

since $\int_{\mathbb{G}} x(h^{-1} \circ g)\,\mathrm{d}\,g = \int_{\mathbb{G}} x(g)\,\mathrm{d}\,g$ for any $x \in L_2(\mathbb{G})$ and $h \in \mathbb{G}$.

CHAPTER 3

THE FAST SLEPIAN TRANSFORM

As explained in Section 2.7, the DPSS basis $\boldsymbol{S}_K$ (a.k.a the Slepian basis, defined in Section 2.6) is the best basis (in a MSE sense) of K columns to represent the discrete vector one obtains when collecting a finite set of uniform samples from a baseband analog signal, the motivating example of parameterized subspace model listed in Section 1.2. Thus, the DPSS basis forms the foundation for the analysis of the parameterized subspace model listed in Section 1.2. In this chapter[3], we present a fast method for computing approximate projections and transforms onto the leading DPSS vectors. The complexity of the resulting algorithms is comparable to the FFT, and scales favorably as the quality of the desired approximation is increased. In the process of bounding the complexity of these algorithms, we also establish new nonasymptotic results on the eigenvalue distribution of discrete time-frequency localization operators. We also establish new nonasymptotic results on the eigenvalue distribution of periodic discrete time-frequency localization operators.

3.1 The Eigenvalue Distribution of Discrete Time-Frequency Localization Operators

We begin with non-asymptotic results on the distribution of the DPSS eigenvalues. Comparing (2.16) with (2.18) we see that the prolate matrix $\boldsymbol{B}_{N,W}$ and the orthogonal projector onto the partial Fourier matrix $\boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ share a somewhat similar structure, where $\boldsymbol{B}_{N,W}$ is a Toeplitz matrix with rows (or columns) given by the sinc function, whereas $\boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ is a circulant matrix with rows (or columns) given by the digital sinc or Dirichlet function. In Theorem 3.1, which is proven in Section A.1, we show that up to a small approximation error ϵ, the difference between these two matrices has a rank of $O(\log N \log \frac{1}{\epsilon})$.

[3]This work is in collaboration with Mark A. Davenport, Santhosh Karnik, Justin Romberg, and Michael B. Wakin [74, 75, 143].

Theorem 3.1. *Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. Then for any $\epsilon \in (0, \frac{1}{2})$, there exist $N \times R_L$ matrices $\boldsymbol{L}_1$,$\boldsymbol{L}_2$ and an $N \times N$ matrix $\boldsymbol{E}_F$ such that*

$$\boldsymbol{B}_{N,W} = \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^* + \boldsymbol{L}_1\boldsymbol{L}_2^* + \boldsymbol{E}_F,$$

where

$$R_L \leq C_N \log\left(\frac{15}{\epsilon}\right), \quad C_N = \left(\frac{4}{\pi^2}\log(8N) + 6\right), \quad \textit{and} \quad \|\boldsymbol{E}_F\| \leq \epsilon.$$

We also note that the proof of Theorem 3.1 provides an explicit construction of the matrices $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$, which could be of use in practice.

An important consequence of Theorem 3.1 which will be useful to us, and which is also of independent interest, is that it can be used to establish a nonasymptotic bound on the number of eigenvalues $\lambda_{N,W}^{(\ell)}$ of $\boldsymbol{B}_{N,W}$ in the "transition region" between ϵ and $1-\epsilon$. In particular, Lemma 2.1 tells as that in the limit as $N \to \infty$ we will have that the first $\approx 2NW$ eigenvalues will approach 1 while the last $\approx N(1-2W)$ eigenvalues will approach 0. However, this does not address precisely how many eigenvalues we can expect to find between ϵ and $1-\epsilon$.

In [118], it is shown that for any $b \in \mathbb{R}$, if $k = \lfloor 2WN + \frac{b}{\pi}\log N \rfloor$, then $\lambda_{N,W}^{(k)} \to (1+e^{\pi b})^{-1}$ as $N \to \infty$. By setting $b = \frac{1}{\pi}\log(\frac{1}{\epsilon} - 1)$, we get $\lambda_{N,W}^{(k)} \to \epsilon$. Similarly, by setting $b = -\frac{1}{\pi}\log(\frac{1}{\epsilon} - 1)$, we get $\lambda_{N,W}^{(k)} \to 1-\epsilon$. This gives us the asymptotic result:

$$\#\{\ell : \epsilon \leq \lambda_{N,W}^{(\ell)} \leq 1-\epsilon\} \sim \frac{2}{\pi^2}\log N \log\left(\frac{1}{\epsilon} - 1\right). \tag{3.1}$$

Figure 3.1 shows a numerical comparison of $\#\{\ell : \epsilon \leq \lambda_{N,W}^{(\ell)} \leq 1-\epsilon\}$ and $\frac{2}{\pi^2}\log N \log\left(\frac{1}{\epsilon} - 1\right)$.

A nonasymptotic bound on the width of this transition region is given in [148], which shows that for any $N \in \mathbb{N}$, $W \in (0, \frac{1}{2})$, and $\epsilon \in (0, \frac{1}{2})$,

$$\#\{\ell : \epsilon \leq \lambda_{N,W}^{(\ell)} \leq 1-\epsilon\} \leq \frac{\frac{2}{\pi^2}\log(N-1) + \frac{2}{\pi^2}\frac{2N-1}{N-1}}{\epsilon(1-\epsilon)}.$$

This bound correctly highlights the logarithmic dependence on N, but can be quite poor when ϵ is very small ($O(1/\epsilon)$ as opposed to the $O(\log(1/\epsilon))$ dependence in the asymptotic

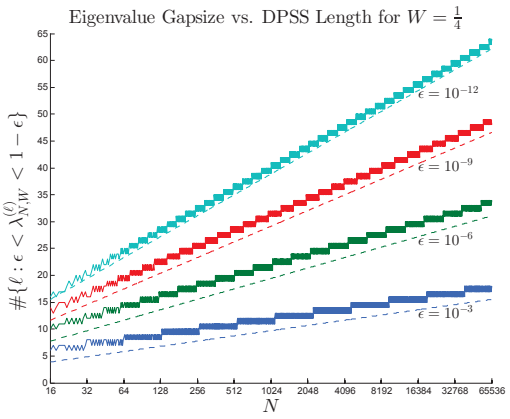


Figure 3.1: The solid lines represent the size of eigenvalue gap for $2^4 \leq N \leq 2^{16}$, $W = \frac{1}{4}$, and $\epsilon = 10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}$. The dashed lines represent the asymptotic result in (3.1). Note that the size of the eigenvalue gap appears to grow linearly with $\log N$ and linearly with $\log(\frac{1}{\epsilon} - 1)$.

result). In the following corollary of Theorem 3.1, we significantly sharpen this bound in terms of its dependence on ϵ to within a constant factor of the optimal asymptotic result. The intuition behind this result is that Theorem 3.1 demonstrates that $\boldsymbol{B}_{N,W}$ can be approximated as $\boldsymbol{F}_{N,W}\boldsymbol{F}^*_{N,W}$ (a matrix whose eigenvalues are all either equal to 1 or 0) plus a low-rank correction, and the rank of this correction limits the number of possible eigenvalues in the transition region.

Corollary 3.1. *Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. Then for any $\epsilon \in (0, \frac{1}{2})$,*

$$\#\{\ell : \epsilon < \lambda^{(\ell)}_{N,W} < 1 - \epsilon\} \leq 2C_N \log\left(\frac{15}{\epsilon}\right).$$

Here C_N is the constant (depending on $\log(N)$) specified in Theorem 3.1.

This result is analogous to the main result of [69], which recently established similar nonasymptotic results concerning the eigenvalue distribution of the *continuous* time-frequency localization operator. In fact, while our approach is quite a bit different than that of [69], it is also possible to establish a version of Corollary 3.1 (with different constants) using some of the same proof techniques as [69].

Finally, we describe a few additional consequences of these results. Recall that $\boldsymbol{B}_{N,W} = \boldsymbol{S}_{N,W}\boldsymbol{\Lambda}_{N,W}\boldsymbol{S}_{N,W}^*$. From Corollary 3.1 we have that the diagonal entries of the matrix $\boldsymbol{\Lambda}_{N,W}$ are mostly very close to 1 or 0, with only a small number of eigenvalues lying in between. Thus, recalling that $\boldsymbol{S}_K$ denotes the $N \times K$ matrix containing the first K elements of the Slepian basis $\boldsymbol{S}_{N,W}$, it is reasonable to expect that $\boldsymbol{B}_{N,W}$ and $\boldsymbol{S}_K\boldsymbol{S}_K^*$ (the matrix obtained by setting the top K eigenvalues to 1 and the remainder to 0) should be within a low-rank correction when $K \approx 2NW$. The following corollary shows that this is indeed the case.

Corollary 3.2. *Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. For any $\epsilon \in (0, \frac{1}{2})$, fix K to be such that $\lambda_{N,W}^{(K-1)} > \epsilon$ and $\lambda_{N,W}^{(K)} < 1 - \epsilon$. Then there exist $N \times r_2$ matrices $\boldsymbol{U}_1, \boldsymbol{U}_2$ and an $N \times N$ matrix $\boldsymbol{E}_2$ such that*

$$\boldsymbol{S}_K\boldsymbol{S}_K^* = \boldsymbol{B}_{N,W} + \boldsymbol{U}_1\boldsymbol{U}_2^* + \boldsymbol{E}_2,$$

where

$$r_2 \leq \left(\frac{8}{\pi^2}\log(8N) + 12\right)\log\left(\frac{15}{\epsilon}\right) \quad \textit{and} \quad \|\boldsymbol{E}_2\| \leq \epsilon.$$

Similarly, consider the rank-K truncated peudoinverse of $\boldsymbol{B}_{N,W}$ where $K \approx 2NW$ (which we will denote by $\boldsymbol{B}_{N,W}^\dagger$). Since most of the first K eigenvalues of $\boldsymbol{B}_{N,W}$ are very close to 1, most of the first K eigenvalues of $\boldsymbol{B}_{N,W}^\dagger$ will also be close to 1. Also, most of the last $N - K$ eigenvalues of $\boldsymbol{B}_{N,W}$ are very close to 0, and by definition the last $N - K$ eigenvalues of $\boldsymbol{B}_{N,W}^\dagger$ are exactly 0. Hence, it is reasonable to expect that $\boldsymbol{B}_{N,W}$ and $\boldsymbol{B}_{N,W}^\dagger$ are within a low-rank correction when $K \approx 2NW$. The following corollary shows that this is indeed the case.

Corollary 3.3. *Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. For any $\epsilon \in (0, \frac{1}{2})$, fix K to be such that $\lambda_{N,W}^{(K-1)} > \epsilon$ and $\lambda_{N,W}^{(K)} < 1 - \epsilon$, and let $\boldsymbol{B}_{N,W}^\dagger$ be the rank-K truncated pseudoinverse of $\boldsymbol{B}_{N,W}$. Then there exist $N \times r_3$ matrices $\boldsymbol{U}_3, \boldsymbol{U}_4$ and an $N \times N$ matrix $\boldsymbol{E}_3$ such that*

$$\boldsymbol{B}_{N,W}^\dagger = \boldsymbol{B}_{N,W} + \boldsymbol{U}_3\boldsymbol{U}_4^* + \boldsymbol{E}_3,$$

where

$$r_3 \leq \left(\frac{8}{\pi^2}\log(8N) + 12\right)\log\left(\frac{15}{\epsilon}\right) \quad \textit{and} \quad \|\boldsymbol{E}_3\| \leq 3\epsilon.$$

A similar decomposition of the pseudo-inverse of $\boldsymbol{B}_{N,W}$, also based on a partial Fourier tranform plus a low rank update, was presented in [95]. Our result above gives an explicit non-asymptotic bound on the rank of the update required to achieve a certain accuracy.

Also, consider the matrix $\boldsymbol{B}_{N,W}^{(\text{tik})} = (\boldsymbol{B}_{N,W}^2 + \alpha\mathbf{I})^{-1}\boldsymbol{B}_{N,W}$ where $\alpha > 0$. (Note that this matrix is associated with Tikhonov regularization, i.e. for a given $\boldsymbol{y} \in \mathbb{C}^N$, the vector $\boldsymbol{x} \in \mathbb{C}^N$ which minimizes $\|\boldsymbol{y} - \boldsymbol{B}_{N,W}\boldsymbol{x}\|_2^2 + \alpha\|\boldsymbol{x}\|_2^2$ is given by $\boldsymbol{x} = (\boldsymbol{B}_{N,W}^2 + \alpha\mathbf{I})^{-1}\boldsymbol{B}_{N,W}\boldsymbol{y} = \boldsymbol{B}_{N,W}^{(\text{tik})}\boldsymbol{y}$). Since most of the eigenvalues of $\boldsymbol{B}_{N,W}$ are either very close to 1 or very close to 0, most of the eigenvalues of $\boldsymbol{B}_{N,W}^{(\text{tik})}$ are either very close to $\frac{1}{1+\alpha}$ or very close to 0. Hence, it is reasonable to expect that $\frac{1}{1+\alpha}\boldsymbol{B}_{N,W}$ and $\boldsymbol{B}_{N,W}^{(\text{tik})}$ are within a low-rank correction. The following corollary shows that this is indeed the case.

Corollary 3.4. *Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ and $\alpha > 0$ be given, and define $\boldsymbol{B}_{N,W}^{(tik)} = (\boldsymbol{B}_{N,W}^2 + \alpha\mathbf{I})^{-1}\boldsymbol{B}_{N,W}$. Then, for any $\epsilon \in (0, \frac{1}{2})$, there exists an $N \times r_4$ matrix $\boldsymbol{U}_5$ and an $N \times N$ matrix $\boldsymbol{E}_4$ such that*

$$\boldsymbol{B}_{N,W}^{(tik)} = \frac{1}{1+\alpha}\boldsymbol{B}_{N,W} + \boldsymbol{U}_5\boldsymbol{U}_5^* + \boldsymbol{E}_4,$$

where

$$r_4 \leq \left(\frac{8}{\pi^2}\log(8N) + 12\right)\log\left(\frac{15}{\min(\alpha(1+\alpha)\epsilon, \frac{1}{3}\epsilon)}\right) \quad \textit{and} \quad \|\boldsymbol{E}_4\| \leq \epsilon.$$

In Section 3.2, we will use Theorem 3.1 along with Corollaries 3.2, 3.3, and 3.4 to derive fast algorithms for working with the Slepian basis.

3.2 The Fast Slepian Transform

A fast factorization of $\boldsymbol{S}_K\boldsymbol{S}_K^*$

Suppose we wish to compress a vector $\boldsymbol{x} \in \mathbb{C}^N$ of N uniformly spaced samples of a signal down to a vector of $K \approx 2NW$ elements in such a way that best preserves the DTFT of the signal over $|f| \leq W$. We can do this by storing $\boldsymbol{S}_K^*\boldsymbol{x}$, which is a vector of $K < N$

elements, and then later recovering $\boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x}$, which contains nearly all of the energy of the signal in the frequency band $|f| \leq W$. However, naïve multiplication of $\boldsymbol{S}_K$ or $\boldsymbol{S}_K^*$ takes $O(NK) = O(2WN^2)$ operations. For certain applications, this may be intractable.

If we combine the results of Corollary 3.2 along with that of Theorem 3.1, we get that

$$\begin{aligned}\boldsymbol{S}_K \boldsymbol{S}_K &= \boldsymbol{B}_{N,W} + \boldsymbol{U}_1 \boldsymbol{U}_2^* + \boldsymbol{E}_2 \\ &= \boldsymbol{F}_{N,W} \boldsymbol{F}_{N,W}^* + \boldsymbol{L}_1 \boldsymbol{L}_2^* + \boldsymbol{U}_1 \boldsymbol{U}_2^* + \boldsymbol{E}_1 + \boldsymbol{E}_2 \\ &= \boldsymbol{T}_1 \boldsymbol{T}_2^* + \boldsymbol{E}_1 + \boldsymbol{E}_2\end{aligned}$$

where

$$\boldsymbol{T}_1 = \begin{bmatrix} \boldsymbol{F}_{N,W} & \boldsymbol{L}_1 & \boldsymbol{U}_1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{T}_2 = \begin{bmatrix} \boldsymbol{F}_{N,W} & \boldsymbol{L}_2 & \boldsymbol{U}_2 \end{bmatrix}.$$

Both $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are $N \times K'$ matrices where

$$K' = 2NW' + r_1 + r_2 \leq \lceil 2NW \rceil + \left(\frac{12}{\pi^2} \log(8N) + 18 \right) \log\left(\frac{15}{\epsilon} \right).$$

So we can compress $\boldsymbol{x}$ by computing $\boldsymbol{T}_2^* \boldsymbol{x}$, which is a vector of $K' \approx 2NW$ elements, and then later recover $\boldsymbol{T}_1 \boldsymbol{T}_2^* \boldsymbol{x}$. By using the triangle inequality, we have $\| \boldsymbol{S}_K \boldsymbol{S}_K^* - \boldsymbol{T}_1 \boldsymbol{T}_2^* \| = \| \boldsymbol{E}_1 + \boldsymbol{E}_2 \| \leq \| \boldsymbol{E}_1 \| + \| \boldsymbol{E}_2 \| \leq 2\epsilon$. Hence, $\| \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x} - \boldsymbol{T}_1 \boldsymbol{T}_2^* \boldsymbol{x} \|_2 \leq 2\epsilon \| \boldsymbol{x} \|_2$ for any vector $\boldsymbol{x} \in \mathbb{C}^N$. Both $\boldsymbol{F}_{N,W}$ and $\boldsymbol{F}_{N,W}^*$ can be applied to a vector in $O(N \log N)$ operations via the FFT. Since $\boldsymbol{L}_1$, $\boldsymbol{L}_2$, $\boldsymbol{U}_1$, and $\boldsymbol{U}_2$ are $N \times O(\log N \log \frac{1}{\epsilon})$ matrices, $\boldsymbol{L}_1$, $\boldsymbol{L}_2^*$, $\boldsymbol{U}_1$, and $\boldsymbol{U}_2^*$ can each be applied to a vector in $O(N \log N \log \frac{1}{\epsilon})$ operations. Therefore, computing $\boldsymbol{T}_2^* \boldsymbol{x}$ and later recovering $\boldsymbol{T}_1 \boldsymbol{T}_2^* \boldsymbol{x}$ (as an approximation for $\boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x}$) takes $O(N \log N \log \frac{1}{\epsilon})$ operations.

Fast projections onto the range of $\boldsymbol{S}_K$

Alternatively, if we only require computing the projected vector $\boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x}$, and compression is not required, then there is a simpler solution. Corollary 3.2 tells us that $\| \boldsymbol{S}_K \boldsymbol{S}_K^* - (\boldsymbol{B}_{N,W} + \boldsymbol{U}_1 \boldsymbol{U}_2^*) \| \leq \epsilon$, and thus, $\| \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x} - (\boldsymbol{B}_{N,W} \boldsymbol{x} + \boldsymbol{U}_1 \boldsymbol{U}_2^* \boldsymbol{x}) \|_2 \leq \epsilon \| \boldsymbol{x} \|_2$ for any vector $\boldsymbol{x} \in \mathbb{C}^N$. Since $\boldsymbol{B}_{N,W}$ is a Toeplitz matrix, computing $\boldsymbol{B}_{N,W} \boldsymbol{x}$ can be done in $O(N \log N)$ operations via the FFT. Since $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are $N \times O(N \log N \log \frac{1}{\epsilon})$ matrices, computing $\boldsymbol{U}_1 \boldsymbol{U}_2^* \boldsymbol{x}$ can

be done in $O(N \log N \log \frac{1}{\epsilon})$ operations. Therefore, we can compute $\boldsymbol{B}_{N,W}\boldsymbol{x} + \boldsymbol{U}_1\boldsymbol{U}_2^*\boldsymbol{x}$ as an approximation to $\boldsymbol{S}_K\boldsymbol{S}_K^*\boldsymbol{x}$ using only $O(N \log N \log \frac{1}{\epsilon})$ operations.

Fast rank-K truncated pseudoinverse of $\boldsymbol{B}_{N,W}$

A closely related problem to working with the matrix $\boldsymbol{S}_K\boldsymbol{S}_K^*$ concerns the task of solving a linear system of the form $\boldsymbol{y} = \boldsymbol{B}_{N,W}\boldsymbol{x}$. Since the prolate matrix has several eigenvalues that are close to 0, the system is often solved by using the rank-K truncated pseudoinverse of $\boldsymbol{B}_{N,W}$ where $K \approx 2NW$. Even if the pseudoinverse is precomputed and factored ahead of time, it still takes $O(NK) = O(2WN^2)$ operations to apply to a vector $\boldsymbol{y} \in \mathbb{C}^N$. Corollary 3.3 tells us that $\|\boldsymbol{B}_{N,W}^\dagger - (\boldsymbol{B}_{N,W} + \boldsymbol{U}_3\boldsymbol{U}_4^*)\| \leq 3\epsilon$, and thus, $\|\boldsymbol{B}_{N,W}^\dagger\boldsymbol{y} - (\boldsymbol{B}_{N,W}\boldsymbol{y} + \boldsymbol{U}_3\boldsymbol{U}_4^*\boldsymbol{y})\|_2 \leq 3\epsilon\|\boldsymbol{y}\|_2$ for any vector $\boldsymbol{y} \in \mathbb{C}^N$. Again, computing $\boldsymbol{B}_{N,W}\boldsymbol{y}$ can be done in $O(N \log N)$ operations using the FFT, and computing $\boldsymbol{U}_3\boldsymbol{U}_4^*\boldsymbol{y}$ can be done in $O(N \log N \log \frac{1}{\epsilon})$ operations. Therefore, we can compute $\boldsymbol{B}_{N,W}\boldsymbol{y} + \boldsymbol{U}_3\boldsymbol{U}_4^*\boldsymbol{y}$ as an approximation to $\boldsymbol{B}_{N,W}^\dagger\boldsymbol{y}$ using only $O(N \log N \log \frac{1}{\epsilon})$ operations.

Fast Tikhonov regularization involving $\boldsymbol{B}_{N,W}$

Another approach to solving the ill-conditioned system $\boldsymbol{y} = \boldsymbol{B}_{N,W}\boldsymbol{x}$ is to use Tikhonov regularization, i.e., minimize $\|\boldsymbol{y} - \boldsymbol{B}_{N,W}\boldsymbol{x}\|_2^2 + \alpha\|\boldsymbol{x}\|_2^2$ where $\alpha > 0$ is a regularization parameter. The solution to this minimization problem is $\boldsymbol{x} = (\boldsymbol{B}_{N,W}^2 + \alpha\mathbf{I})^{-1}\boldsymbol{B}_{N,W}\boldsymbol{y}$. Even if the matrix $\boldsymbol{B}_{N,W}^{(\text{tik})} = (\boldsymbol{B}_{N,W}^2 + \alpha\mathbf{I})^{-1}\boldsymbol{B}_{N,W}$ is computed ahead of time, it still takes $O(N^2)$ operations to apply to a vector $\boldsymbol{y}$. Corollary 3.4 tells us that $\|\boldsymbol{B}_{N,W}^{(\text{tik})} - (\boldsymbol{B}_{N,W} + \boldsymbol{U}_5\boldsymbol{U}_5^*)\| \leq \epsilon$, and thus, $\|\boldsymbol{B}_{N,W}^{(\text{tik})}\boldsymbol{y} - (\boldsymbol{B}_{N,W}\boldsymbol{y} + \boldsymbol{U}_5\boldsymbol{U}_5^*\boldsymbol{y})\|_2 \leq \epsilon\|\boldsymbol{y}\|_2$ for any vector $\boldsymbol{y} \in \mathbb{C}^N$. Again, computing $\boldsymbol{B}_{N,W}\boldsymbol{y}$ can be done in $O(N \log N)$ operations via the FFT. Since $\boldsymbol{U}_5$ has size $N \times O\left(\log N \max(\log \frac{1}{\alpha\epsilon}, \log \frac{1}{\epsilon})\right)$, computing $\boldsymbol{U}_5\boldsymbol{U}_5^*\boldsymbol{y}$ can be done in $O(N \log N \max(\log \frac{1}{\alpha\epsilon}, \log \frac{1}{\epsilon}))$ operations. Therefore, we can compute $\boldsymbol{B}_{N,W}\boldsymbol{y} + \boldsymbol{U}_5\boldsymbol{U}_5^*\boldsymbol{y}$ as an approximation to $\boldsymbol{B}_{N,W}^{(\text{tik})}\boldsymbol{y}$ using only $O\left(N \log N \max(\log \frac{1}{\alpha\epsilon}, \log \frac{1}{\epsilon})\right)$ operations.

The least-squares problems above involve the inverse of $\boldsymbol{B}_{N,W}$, a symmetric semi-definite Toeplitz matrix. There is a long history of "superfast" algorithms for inverting such systems

in the signal processing [98, 72] and numerical linear algebra [4, 62, 8] literature. These algorithms take a number of different forms. They usually work by breaking the matrix into smaller blocks, either hierarchically [94] or recusively [9, 102], and then exploiting the structure of the matrix to efficiently combine the solutions of smaller systems into a solution for the entire system. The overall computational complexity of these algorithms is $O(N \log^2 N)$ for the first solve with a given matrix, and $O(N \log N)$ for subsequent solves. An overview of these methods can be found in [129].

The approach suggested by Corollary 3.3 (and the regularized version in Corollary 3.4) have the same run time of $O(N \log N)$, but are based on entirely different principles. Theorem 3.1 essentially states that the matrix $\boldsymbol{B}_{N,W}$ is a low-rank update away from an orthoprojection, and this orthoprojection can be computed quickly using the FFT. Corollaries 3.3 and 3.4 show that this property also holds for the (regularized) pseudo-inverse. These mathematical results show that this particular system can be very closely approximated by a sum of circulant and low-rank matrices, which leads directly to efficient algorithms for solving least-squares problems.

3.3 Applications

Owing to the concentration in the time and frequency domains, the Slepian basis vectors have proved to be useful in numerous signal processing problems [2, 32, 118, 140, 31]. Linear systems of equations involving the prolate matrix $\boldsymbol{B}_{N,W}$ also arise in several problems, such as band-limited extrapolation [118]. In this section, we describe some specific applications that stand to benefit from the fast constructions described above.

i. **Representation and compression of sampled bandlimited and multiband signals.** Consider a length-N vector $\boldsymbol{x}$ obtained by uniformly sampling a baseband analog signal $x(t)$ over the time interval $[0, NT_s)$ with sampling period $T_s \leq \frac{1}{B_{\text{band}}}$ chosen to satisfy the Nyquist sampling rate. Here, $x(t)$ is assumed to be bandlimited with frequency range $[-B_{\text{band}}/2, B_{\text{band}}/2]$. Under this assumption, the sample vector $\boldsymbol{x}$ can be expressed as

$$\boldsymbol{x}[n] = \int_{-W}^{W} X(f) e^{j2\pi f n}\, df,\ n = 0, 1, \ldots, N-1, \tag{3.2}$$

or equivalently,

$$\boldsymbol{x} = \int_{-W}^{W} X(f) \boldsymbol{e}_f\, df \tag{3.3}$$

where $W = T_s B_{\text{band}}/2 \leq \frac{1}{2}$ and $X(f)$ is the DTFT of the infinite sample sequence $x[n] = x(nT_s), n \in \mathbb{Z}$. Such finite-length vectors of samples from bandlimited signals arise in problems such as time-variant channel estimation [140] and mitigation of narrowband interference [29]. Solutions to these and many other problems benefit from representations that efficiently capture the structure inherent in vectors $\boldsymbol{x}$ of the form (3.3).

In [32], the authors showed that such a vector $\boldsymbol{x}$ has a low-dimensional structure by building a dictionary in which $\boldsymbol{x}$ can be approximated with a small number of atoms. The $N \times N$ DFT basis is insufficient to capture the low dimensional structure in $\boldsymbol{x}$ due to the "DFT leakage" phenomenon. In particular, the DFT basis is comprised of vectors $\boldsymbol{e}_f$ with f sampled uniformly between $-1/2$ and $1/2$. From (3.3), one can interpret $\boldsymbol{x}$ as being comprised of a linear combination of vectors $\boldsymbol{e}_f$ with f ranging continuously between $-W$ and W. It is natural to ask whether $\boldsymbol{x}$ could be efficiently represented using only the DFT vectors $\boldsymbol{e}_f$ with f between $-W$ and W; in particular, these are the columns of the matrix $\boldsymbol{F}_{N,W}$ defined in (2.17). Unfortunately, this is not the case—while a majority of the energy of $\boldsymbol{x}$ can be captured using the columns of $\boldsymbol{F}_{N,W}$, a nontrivial amount will be missed and this is contained in the familiar sidelobes in the DFT outside the band of interest.

An efficient alternative to the partial DFT $\boldsymbol{F}_{N,W}$ is given by the partial Slepian basis $\boldsymbol{S}_K$ when $K \approx 2NW$. In [32], for example, it is established that when $\boldsymbol{x}$ is generated by sampling a bandlimited analog random process with flat power spectrum over $[-B_{\text{band}}/2, B_{\text{band}}/2]$, and when one chooses $K = 2NW(1+\epsilon)$, then on average $\boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x}$ will capture all but an exponentially small amount of the energy from $\boldsymbol{x}$. Zemen and Mecklenbräuke [140] showed that expressing the time-varying subcarrier coefficients in a Slepian basis yields better

performance than that obtained with a DFT basis, which suffers from frequency leakage.

By modulating the (baseband) Slepian basis vectors to different frequency bands and then merging these dictionaries, one can also obtain a new dictionary that offers an efficient representation of sampled *multiband* signals. Zemen et al. [141] proposed one such dictionary for estimating a time-variant flat-fading channel whose spectral support is a union of several intervals. In the context of compressive sensing, Davenport and Wakin [32] investigated multiband modulated DPSS dictionaries for sparse recovery of sampled multiband signals, and Sejdić et al. [110] applied such dictionaries for recovery of physiological signals from compressive measurements. In Chapter 5, we employed such dictionaries for detecting targets behind the wall in through-the-wall radar imaging, and modulated DPSS's can also be useful for mitigating wall clutter [2].

In summary, many of the above mentioned problems are facilitated by projecting a length-N vector onto the subspace spanned by the first $K \approx 2NW$ Slepian basis vectors (i.e., computing $\boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{x}$). One version of the Block-Based CoSaMP algorithm in [32] involves computing the projection of a vector onto the column space of a modulated DPSS dictionary. The channel estimates proposed in [111] are based on the projection of the subcarrier coefficients onto the column space of the modulated multiband DPSS dictionary. Of course, one can also compress $\boldsymbol{x}$ by keeping the $\approx 2NW$ Slepian basis coefficients $\boldsymbol{S}_K^* \boldsymbol{x}$ instead of the N entries of $\boldsymbol{x}$. Computationally, all of these problems benefit from having a fast Slepian transform: whereas direct matrix-vector multiplication would require $O(2NW \cdot N) = O(N^2)$ operations, the fast Slepian constructions allow these computations to be approximated in only $O\left(N \log N \log \frac{1}{\epsilon}\right)$ operations.

ii. **Prolate matrix linear systems.** Linear equations of the form $\boldsymbol{B}_{N,W} \boldsymbol{y} = \boldsymbol{b}$ arise naturally in signal processing. For example, suppose we obtain the length-N sampled bandlimited vector $\boldsymbol{x}$ as defined in (3.2) and we are interested in estimating the infinite-length sequence $x[n] = x(nT_s),\ \forall n \in \mathbb{Z}$. The discrete-time signal $x[n]$ is assumed to be bandlimited to $[-W, W]$ for $W < \frac{1}{2}$. Let $\mathcal{I}_N : \ell_2(\mathbb{Z}) \to \mathbb{C}^N$ denote the index-limiting operator

that restricts a sequence to its entries on $[N]$ (and produces a vector of length N); that is $\mathcal{I}_N(y)[m] := y[m]$ for all $m \in \{0, 1, \ldots, N-1\}$. Also, recall that $\mathcal{B}_W : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ denotes the band-limiting operator that bandlimits the DTFT of a discrete-time signal to the frequency range $[-W, W]$. Given $\boldsymbol{x}$, the least-squares estimate $\widehat{x}[n] \in l_2(\mathbb{Z})$ for the infinite-length bandlimited sequence takes the form

$$\widehat{x}[n] = [(\mathcal{I}_N \mathcal{B}_W)^\dagger \boldsymbol{x}][n] = \sum_{m=0}^{N-1} v[m] \frac{\sin 2\pi W(n-m)}{\pi(n-m)},$$

where $\boldsymbol{v} = \boldsymbol{B}_{N,W}^\dagger \boldsymbol{x}$.

Another problem involves linear prediction of bandlimited signals based on past samples. Suppose $x(t)$ is a continuous, zero-mean, wide sense stationary random process with power spectrum

$$P_x(F) = \begin{cases} \frac{1}{B_{\text{band}}}, & -\frac{B_{\text{band}}}{2} \leq F \leq \frac{B_{\text{band}}}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $x[n] = x(nT_s)$ denote the samples acquired from $x(t)$ with a sampling interval of $T_s \leq \frac{1}{B_{\text{band}}}$. A linear prediction of $x[N]$ based on the previous N samples $x[0], x[1], \ldots, x[N-1]$ takes the form [118]

$$\widehat{x}[N] = \sum_{n=0}^{N-1} a_n x[n].$$

Choosing a_n such that $\widehat{x}[N]$ has the minimum mean-squared error is equivalent to solving

$$\min_{a_n} \varrho := \mathbb{E}\left[\left(\sum_{n=0}^{N-1} a_n x[n] - x[N] \right)^2 \right].$$

Let $W = \frac{T_s}{2B_{\text{band}}}$. Taking the derivative of ϱ and setting it to zero yields

$$\boldsymbol{B}_{N,W} \boldsymbol{a} = \boldsymbol{b}$$

with $\boldsymbol{a} = [a_0 \ a_1 \ \cdots \ a_{N-1}]^T$ and $\boldsymbol{b} = \left[\frac{\sin(2\pi WN)}{\pi N} \ \frac{\sin(2\pi W(N-1))}{\pi(N-1)} \ \cdots \ \frac{\sin(2\pi W1)}{\pi 1} \right]^T$. Thus the optimal $\widehat{\boldsymbol{a}}$ is simply given by $\widehat{\boldsymbol{a}} = \boldsymbol{B}_{N,W}^\dagger \boldsymbol{b}$.

We present one more example: the Fourier extension [68]. The partial Fourier series sum

$$y_{N'}(t) = \frac{1}{\sqrt{2}} \sum_{|n| \leq N'} \widehat{y}_n e^{jn\pi t}, \quad \widehat{y}_n = \frac{1}{\sqrt{2}} \int_{-1}^{1} y(t) e^{-jn\pi t} dt$$

of a non-periodic function $y \in L^2_{[-1,1]}$ (such as $y(t) = t$) suffers from the Gibbs phenomenon. One approach to overcome the Gibbs phenomenon is to extend the function y to a function g that is periodic on a larger interval $[-T, T]$ with $T > 1$ and compute the partial Fourier series of g [68]. Let $G_{N''}$ be the space of bandlimited $2T$-periodic functions

$$G_{N''} := \left\{ g : g(t) = \sum_{n=-N''}^{N''} \widehat{g}_n e^{\frac{jn\pi t}{T}}, \ \widehat{g}_n \in \mathbb{C} \right\}.$$

The Fourier extension problem involves finding

$$g_{N''} := \arg \min_{g \in G_{N''}} \|y - g\|_{L^2_{[-1,1]}}. \tag{3.4}$$

The solution $g_{N''}$ is called the Fourier extension of y to the interval $[-T, T]$. Let $\widehat{\boldsymbol{g}} = [\widehat{g}_{-N''} \ \cdots \ \widehat{g}_0 \ \cdots \ \widehat{g}_{N''}]^T$ and define $\mathcal{F}_{N''} : L_2([-1,1]) \to \mathbb{C}^{2N''+1}$ as

$$(\mathcal{F}_{N''}(u))[n] = \frac{1}{\sqrt{2T}} \int_{-1}^{1} u(t) e^{-\frac{jn\pi t}{T}} dt, \quad |n| \leq N''.$$

For convenience, here we index all vectors and matrices beginning at $-N''$. Any minimizer $\widehat{\boldsymbol{g}}$ of the least-squares problem (3.4) must satisfy the normal equations

$$\mathcal{F}_{N''} \mathcal{F}^*_{N''} \widehat{\boldsymbol{g}} = \mathcal{F}_{N''} y, \tag{3.5}$$

where $\mathcal{F}_{N''} y$ can be efficiently approximately computed via the FFT. One can show that $\mathcal{F}_{N''} \mathcal{F}^*_{N''} = \boldsymbol{B}_{N,W}$, where $N = 2N'' + 1$ and $W = \frac{1}{2T} \leq \frac{1}{2}$.

Each of the above least-squares problems could be solved by computing a rank-K truncated pseudo-inverse of $\boldsymbol{B}_{N,W}$ with $K \approx 2NW$. Direct multiplication of this matrix with a vector, however, would require $O(2NW \cdot N) = O(N^2)$ operations. The fast methods we have developed allow a fast approximation to the truncated pseudo-inverse to be applied in only

$O\left(N \log N \log \frac{1}{\epsilon}\right)$ operations.

3.4 Simulations

In this section, we present several numerical simulations comparing our fast, approximate algorithms to the exact versions.

Fast projection onto the span of $\boldsymbol{S}_K$

To test our fast factorization of $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and our fast projection method, we fix the half-bandwidth $W = \frac{1}{4}$ and vary the signal length N over several values between 2^8 and 2^{20}. For each value of N we randomly generate several length-N vectors and project each one onto the span of the first $K = \text{round}(2NW)$ elements of the Slepian basis using the fast factorization $\boldsymbol{T}_1\boldsymbol{T}_2^*$ and the fast projection matrix $\boldsymbol{B}_{N,W} + \boldsymbol{U}_1\boldsymbol{U}_2^*$ for tolerances of $\epsilon = 10^{-3}$, 10^{-6}, 10^{-9}, and 10^{-12}. The prolate matrix, $\boldsymbol{B}_{N,W}$, is applied to the length N vectors via an FFT whose length is the smallest power of 2 that is at least $2N$. For values of $N \leq 12288$, we also projected each vector onto the span of the first K elements of the Slepian basis using the exact projection matrix $\boldsymbol{S}_K\boldsymbol{S}_K^*$. The exact projection could not be tested for values of $N > 12288$ due to computational limitations. A plot of the average time needed to project a vector onto the span of the first $K = \text{round}(2NW)$ elements of the Slepian basis using the exact projection matrix $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and the fast factorization $\boldsymbol{T}_1\boldsymbol{T}_2^*$ is shown in the top left in Figure 3.2. A similar plot comparing the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and the fast projection $\boldsymbol{B}_{N,W} + \boldsymbol{U}_1\boldsymbol{U}_2^*$ is shown in the top right in Figure 3.2. As can be seen in the figures the time required by the exact projection grows quadratically with N, while the time required by the fast factorization as well as the fast projection grows roughly linearly in N.

For the exact projection, all of the first $K = \text{round}(2NW)$ elements of the Slepian basis must be precomputed. For the fast factorization, the low rank matrices $\boldsymbol{L}_1, \boldsymbol{L}_2$ (from Theorem 3.1) and the Slepian basis elements $\boldsymbol{s}_{N,W}^{(\ell)}$ for which $\epsilon < \lambda_{N,W}^{(\ell)} < 1 - \epsilon$ are precomputed. For the fast projection, the FFT of the sinc kernel, as well as the Slepian basis elements $\boldsymbol{s}_{N,W}^{(\ell)}$ for which $\epsilon < \lambda_{N,W}^{(\ell)} < 1 - \epsilon$ are precomputed. A plot of the average precomputation

time needed for both the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ as well as the fast factorization $\boldsymbol{T}_1\boldsymbol{T}_2^*$ is shown in the top left in Figure 3.3. A similar plot comparing the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and the fast projection $\boldsymbol{B}_{N,W} + \boldsymbol{U}_1\boldsymbol{U}_2^*$ is shown in the top right in Figure 3.3. As can be seen in the figures the precomputation time required by the exact projection grows roughly quadratically with N, while the precomputation time required by the fast factorization as well as the fast projection grows just faster than linearly in N.

This experiment was repeated with $W = \frac{1}{16}$ and $W = \frac{1}{64}$ (instead of $W = \frac{1}{4}$). The results for $W = \frac{1}{16}$ and $W = \frac{1}{64}$ are shown in the middle and bottom, respectively, of Figures Figure 3.2 and Figure 3.3. The exact projection onto the first $K \approx 2NW$ elements of the Slepian basis takes $O(NK) = O(2WN^2)$ operations, whereas both our fast factorization and fast projection algorithms take $O(N \log N \log \frac{1}{\epsilon})$ operations. The smaller W gets, the larger N needs to be for our fast methods to be faster than the exact projection via matrix multiplication. If $W \lesssim \frac{1}{N} \log N \log \frac{1}{\epsilon}$, then our fast methods lose their computational advantage over the exact projection. However, in this case the exact projection is fast enough to not require a fast approximate algorithm.

Solving least-squares systems involving $\boldsymbol{B}_{N,W}$

We demonstrate the effectiveness of our fast prolate pseudoinverse method (Corollary 3.3) and our fast prolate Tikhonov regularization method (Corollary 3.4) on an instance of the Fourier extension problem, as described in Section 3.3.

To choose an appropriate function f, we note that if f is continuous and $f(-1) = f(1)$, then the Fourier sum approximations will not suffer from Gibbs phenomenon, and so, there is no need to compute a Fourier extension sum approximation for f. Also, if f is smooth on $[-1, 1]$ but $f(-1) \neq f(1)$, then the Fourier sum approximations will suffer from Gibbs phenomenon, but the Fourier extension series coefficients will decay exponentially fast. Hence, relatively few Fourier extension series coefficients will be needed to accurately approximate f, which makes the least squares problem of solving for these coefficients small enough for our fast methods to not be useful. However, in the case where f is continuous but not smooth

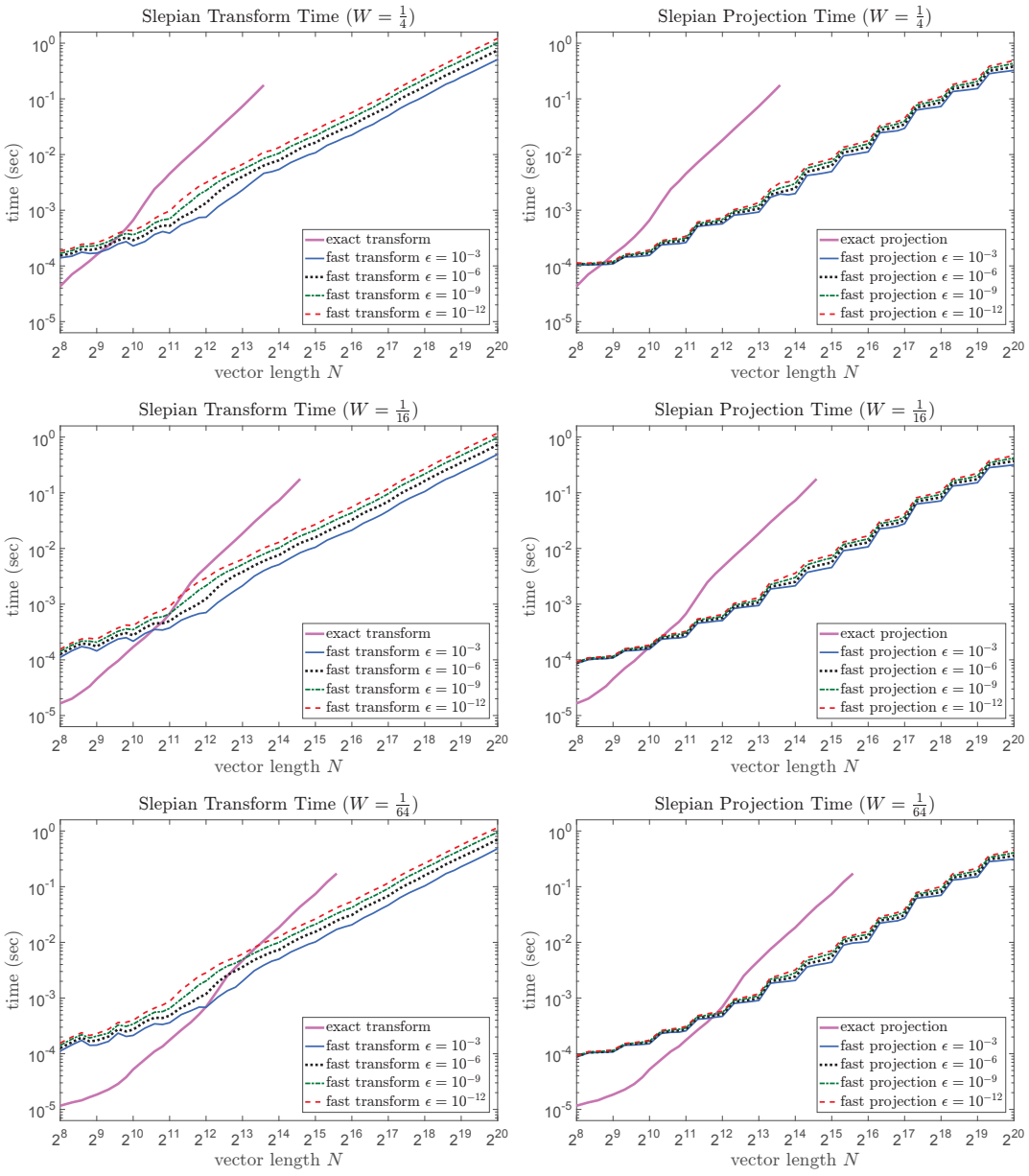


Figure 3.2: (Left) Plots of the average time needed to project a vector onto the first round$(2NW)$ Slepian basis elements using the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and using the fast factorization $\boldsymbol{T}_1\boldsymbol{T}_2^*$. (Right) Plots of the average time needed to project a vector onto the first round$(2NW)$ Slepian basis elements using the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and using the fast projection $\boldsymbol{B}_{N,W} + \boldsymbol{U}_1\boldsymbol{U}_2^*$.

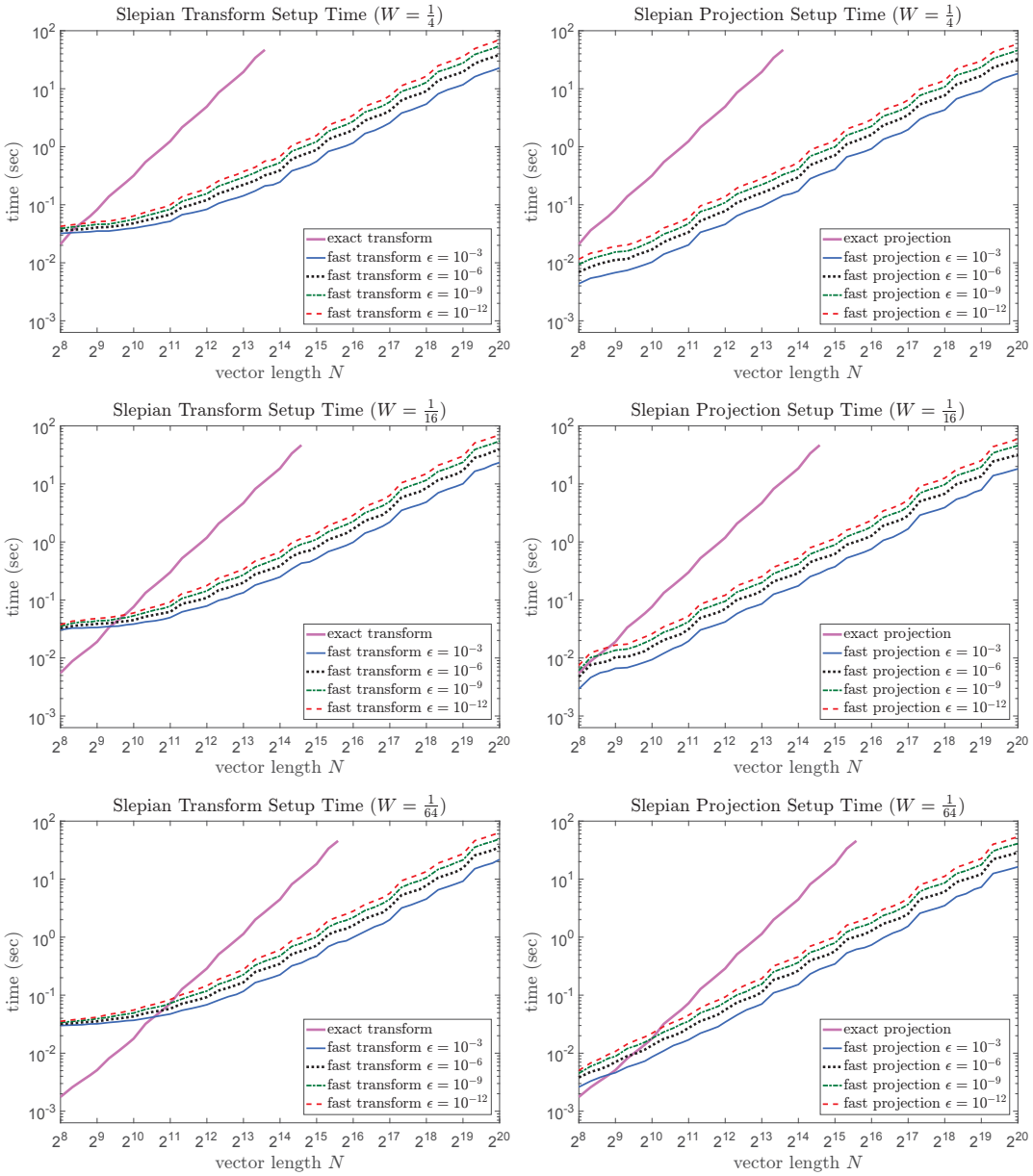


Figure 3.3: (Left) Plots of the average precomputation time for the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and the fast factorization $\boldsymbol{T}_1\boldsymbol{T}_2^*$. (Right) Plots of the average precomputation time for the exact projection $\boldsymbol{S}_K\boldsymbol{S}_K^*$ and the fast projection $\boldsymbol{B}_{N,W} + \boldsymbol{U}_1\boldsymbol{U}_2^*$.

on $[-1,1]$ and $f(-1) \neq f(1)$, the Fourier series will suffer from Gibbs phenomenon, and the Fourier extension series coefficients will decay faster than the Fourier series coefficients, but not exponentially fast. So in this case, the number of Fourier extension series coefficients required to accurately approximate f is not trivially small, but still less than the number of Fourier series coefficients required to accurately approximate f. Hence, computing a Fourier extension sum approximation to f is useful and requires our fast methods.

We construct such a function $f : [-1,1] \to \mathbb{R}$ in the form

$$f(t) = a_0 t + \sum_{\ell=1}^{L} a_\ell \exp(-\tfrac{|t-\mu_\ell|}{\sigma_\ell}) \tag{3.6}$$

where $a_0 = 5$, $L = 500$, and a_ℓ, μ_ℓ, and σ_ℓ are chosen in a random manner. A plot of $f(t)$ over $t \in [-1,1]$ is shown on the left in Figure 3.4. Also on the right in Figure 3.4, we show an example of a Fourier sum approximation and a Fourier extension approximation, both with 401 terms. Notice that the Fourier sum approximation suffers from Gibbs phenomenon near the endpoints of the interval $[-1,1]$, while the Fourier extension approximation does not exhibit such oscillations near the endpoints of $[-1,1]$.

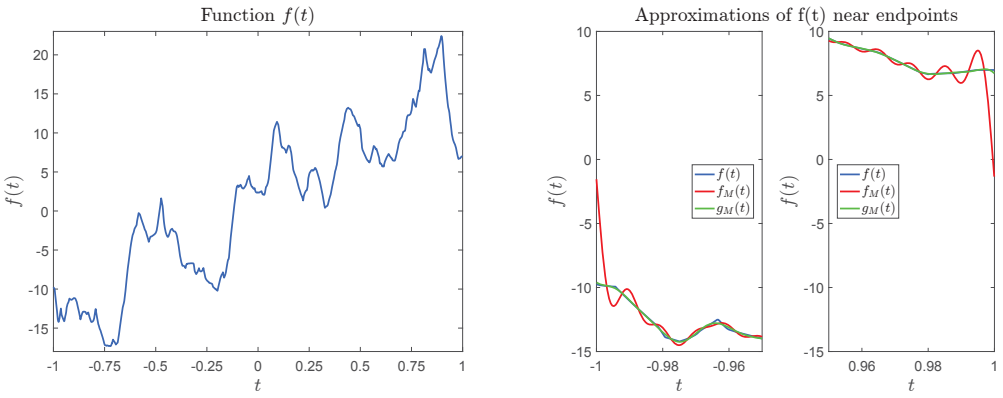


Figure 3.4: (Left) A plot of the function used in the experiments described in (3.6). (Right) Plots of the function, the Fourier sum approximation to $f(t)$ using 401 terms, and the Fourier extension approximation to $f(t)$ using 401 terms. Note that the Fourier sum approximation suffers from Gibbs phenomenon oscillations while the Fourier extension sum does not.

For several positive integers M between 1 and 2560, we compute three approximations to $f(t)$:

1. The $2M+1$ term truncated Fourier series of $f(t)$, i.e.,

$$f_M(t) = \frac{1}{\sqrt{2}} \sum_{m=-M}^{M} \widehat{f}_m e^{j\pi mt}, \text{ where } \widehat{f}_m = \frac{1}{\sqrt{2}} \int_{-1}^{1} f(t) e^{j\pi mt}\, dt.$$

2. The $2M+1$ term Fourier extension of $f(t)$ to the interval $[-T, T]$, i.e.,

$$g_M(t) = \frac{1}{\sqrt{2T}} \sum_{m=-M}^{M} \widehat{g}_m e^{j\pi mt/T},$$

   where $\widehat{y}_m = \frac{1}{\sqrt{2T}} \int_{-1}^{1} f(t) e^{j\pi mt/T}\, dt$ and $\widehat{g} = \boldsymbol{B}^{\dagger}_{(2M+1),\frac{1}{2T}} \widehat{y}$. Here, we pick $T = 1.5$, and we let $\boldsymbol{B}^{\dagger}_{(2M+1),\frac{1}{2T}}$ be the truncated pseudoinverse of $\boldsymbol{B}_{(2M+1),\frac{1}{2T}}$ which zeros out eigenvalues smaller than 10^{-4}.

3. The $2M+1$ term Fourier extension of $f(t)$ to the interval $[-T, T]$ (as described above), except we use the fast prolate pseudoinverse method (Corollary 3.3) with tolerance $\epsilon = 10^{-5}$ instead of the exact truncated pseudoinverse.

The integrals used in computing the coefficients are approximated using an FFT of length 2^{13+q} where $q = \lfloor \log_2 M \rfloor$. By increasing the FFT length with M, we ensure that the coefficients are sufficiently approximated, while also ensuring that the time needed to compute the FFT does not dominate the time needed to solve the system $\boldsymbol{B}_{(2M+1),\frac{1}{2T}} \widehat{g} = \widehat{y}$. Given an approximation $\hat{f}(t)$ to $f(t)$, we quantify the performance via the relative root-mean-square (RMS) error:

$$\frac{\|f - \hat{f}\|_{L^2[-1,1]}}{\|f\|_{L^2[-1,1]}}$$

A plot of the relative RMS error versus M for each of the three approximations to $f(t)$ is shown on the left in Figure 3.5. For values of M at least 448, the Fourier extension $g_M(t)$

(computed with either the exact or the fast pseudoinverse) yielded a relative RMS error at least 10 times lower than that for the truncated Fourier series $f_M(t)$. Using the exact pseudoinverse instead of the fast pseudoinverse does not yield a noticable improvement in the approximation error. A plot of the average time needed to compute the approximation coefficients versus M is shown on the right in Figure 3.5. For large M, computing the Fourier extension coefficients using the fast prolate pseudoinverse is significantly faster than computing the Fourier extension coefficients using the fast prolate pseudoinverse. Also, computing the Fourier extension coefficients using the fast prolate pseudoinverse takes only around twice the time required for computing the Fourier series coefficients.

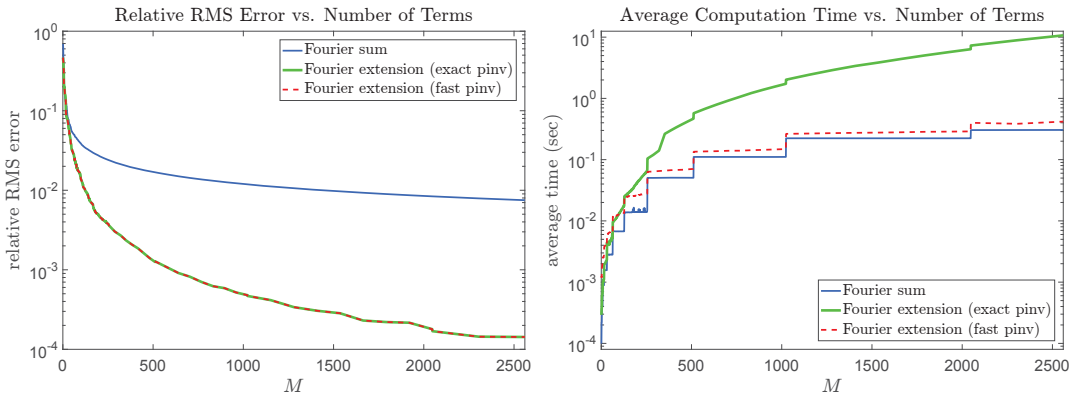


Figure 3.5: A comparison of the relative RMS error (left) and the computation time required (right) for the $2M+1$ term truncated Fourier series as well as the $2M+1$ term Fourier extension using both the exact and fast pseudoinverse methods. Note that the exact and fast methods are virtually indistinguishable in terms of relative RMS error.

We repeated this experiment, except using Tikhonov regularization to solve the system $\boldsymbol{B}_{(2M+1),\frac{1}{2T}}\widehat{g} = \widehat{y}$ instead of the truncated pseudoinverse. We tested both the exact Tikhonov regularization procedure $\widehat{g} = (\boldsymbol{B}^2_{(2M+1),\frac{1}{2T}} + \alpha\mathbf{I})^{-1}\boldsymbol{B}_{(2M+1),\frac{1}{2T}}\widehat{y}$ (for $\alpha = 10^{-8}$) as well as the fast Tikhonov regularization method (Corollary 3.4) with a tolerance of $\epsilon = 10^{-5}$. The results, which are similar to those for the pseudoinverse case, are shown in Figure 3.6.

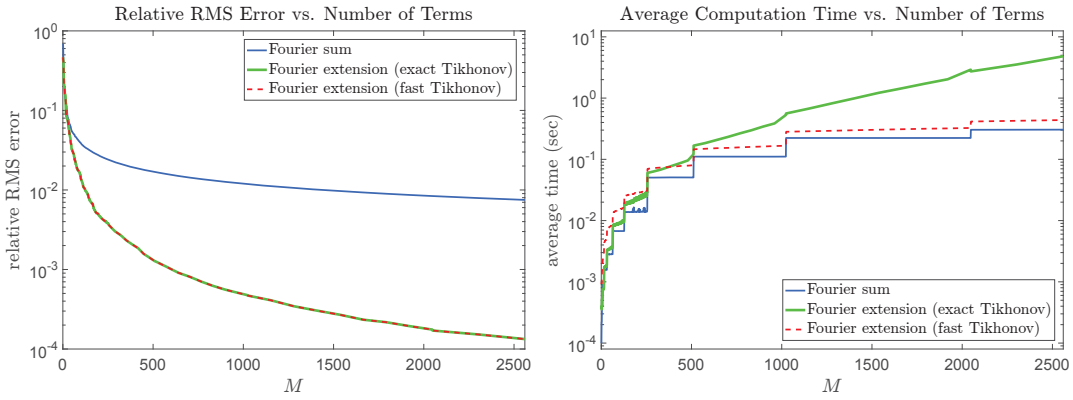


Figure 3.6: A comparison of the relative RMS error (left) and the computation time required (right) for the $2M+1$ term truncated Fourier series as well as the $2M+1$ term Fourier extension using both the exact and fast Tikhonov regularization methods. Note that the exact and fast methods are virtually indistinguishable in terms of relative RMS error.

3.5 The Eigenvalue Distribution of Discrete Periodic Time-Frequency Limiting Operators

The *periodic discrete prolate spheroidal sequences* (PDPSS's), introduced by Jain and Ranganath [71] and Grünbaum [57], are the counterparts of the PSWF's in the finite dimensional case. The PDPSS's are the finite-length vectors whose discrete Fourier transform (DFT) is most concentrated in a given bandwidth. Being simultaneously concentrated in the time and frequency domains makes these vectors useful in a number of signal processing applications. For example, Jain and Ranganath used PDPSS's for extrapolation and spectral estimation of periodic discrete-time signals [71]. PDPSS's were also used for limited-angle reconstruction in tomography [57], for Fourier extension [95], and in [65], the bandpass PDPSS's were used as a numerical approximation to the bandpass PSWF's for studying synchrony in sampled EEG signals.

The distribution of the eigenvalues of a time-frequency limiting operator dictate the (approximate) dimension of the space of signals which are bandlimited and approximately timelimited [120, 118]. Such distributions are known for the case of PSWF's and DPSS's. Specifically, an asymptotic expression for the PSWF eigenvalues was given in [116], and more recently, Israel [69] provided a non-asymptotic bound. Slepian [118] first provided

an asymptotic expression for the DPSS eigenvalues. In [75], we recently provided a non-asymptotic result for the distribution of the DPSS eigenvalues (which improves upon a previous result in [148]).

There exist comparatively few results concerning the PDPSS eigenvalues. In [137], it was shown that unlike the PSWF and DPSS eigenvalues, the PDPSS eigenvalues can exactly achieve 0 and 1 and are degenerate in some cases. A non-asymptotic result on the distribution of the PDPSS eigenvalues was given in [45]. The special distribution of the PDPSS eigenvalues (See Figure 3.7) has been exploited for fast computing Fourier extensions of arbitrary extension length in [95]. In this paper, we provide a finer non-asymptotic result that improves upon the expression in [45]. We also characterize the spectrum of submatrices of the DFT matrix (see Corollary 3.5), which is of independent interest in signal processing. For example, the low rank of DFT submatrices can be exploited for efficiently computing DFT [45].

With abuse of notation[4], let $\mathcal{T}_N : \mathbb{C}^M \to \mathbb{C}^M$ denote a time-limiting operator that only keeps the first $N \leq M$ entries of a vector, i.e., for any $\boldsymbol{x} \in \mathbb{C}^M$,

$$(\mathcal{T}_N(\boldsymbol{x}))[n] := \begin{cases} \boldsymbol{x}[n], & 0 \leq n \leq N-1, \\ 0, & N \leq n \leq M-1. \end{cases}$$

The DFT of any $\boldsymbol{x} \in \mathbb{C}^M$, denoted by $\widehat{\boldsymbol{x}} \in \mathbb{C}^M$, is defined as

$$\widehat{\boldsymbol{x}}[n] := \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \boldsymbol{x}[m] e^{-j\frac{2\pi nm}{M}}, \ n \in [M],$$

where $[M] = \{0, \ldots, M-1\}$. Given $\widehat{\boldsymbol{x}}$, $\boldsymbol{x}$ can be recovered by taking the inverse DFT (IDFT), i.e.,

$$\boldsymbol{x}[m] = \frac{1}{\sqrt{M}} \sum_{n=0}^{M-1} \widehat{\boldsymbol{x}}[n] e^{j\frac{2\pi nm}{M}}, \ m \in [M].$$

Suppose $K \in \mathbb{N}$ such that $2K+1 < M$. Let $\mathcal{B}_K : \mathbb{C}^M \to \mathbb{C}^M$ denote a band-limiting operator that first zeros out the DFT of a vector outside the index range $\mathcal{I}_K := \{0, \ldots, K\} \cup$

[4]In this section, the time- and band-limiting operators are slightly different than the ones in Section 2.5. But it should be clear to distinguish them from the context.

$\{M-K,\ldots,M-1\}$, then returns the corresponding signal in the time domain by taking the IDFT. That is

$$\begin{aligned}(\mathcal{B}_K(\boldsymbol{x}))[m] &:= \frac{1}{\sqrt{M}}\sum_{k\in\mathcal{I}_K}\widehat{\boldsymbol{x}}[k]e^{j\frac{2\pi km}{M}}\\ &= \frac{1}{M}\sum_{k\in\mathcal{I}_K}\sum_{n=0}^{M-1}\boldsymbol{x}[n]e^{-j\frac{2\pi nk}{M}}e^{j\frac{2\pi km}{M}}\\ &= \sum_{n=0}^{M-1}\frac{\sin\left((2K+1)(m-n)\pi/M\right)}{M\sin\left((m-n)\pi/M\right)}\boldsymbol{x}[n].\end{aligned}$$

Denote $W=\frac{2K+1}{2M}<\frac{1}{2}$. Let $\overline{\boldsymbol{B}}_{M,W}\in\mathbb{C}^{M\times M}$ denote a prolate matrix with entries

$$\overline{\boldsymbol{B}}_{M,W}[m,n]=\frac{\sin\left(2\pi W(m-n)\right)}{M\sin\left((m-n)\pi/M\right)},\ m,n\in[M].$$

Note that $\mathcal{B}_K$ is equivalent to $\overline{\boldsymbol{B}}_{M,W}$, whose eigenvectors are given by the PDPSS's [71, 57].

Let $[\overline{\boldsymbol{B}}_{M,W}]_N\in\mathbb{C}^{N\times N}$ be the leading principal submatrix of $\overline{\boldsymbol{B}}_{M,W}$ constructed by removing the last $M-N$ rows and columns from $\overline{\boldsymbol{B}}_{M,W}$. Composing the time- and band-limiting operators, we obtain the linear operator $\mathcal{T}_N\mathcal{B}_K\mathcal{T}_N:\mathbb{C}^M\to\mathbb{C}^M$, which has the same non-zero eigenvalues as $[\overline{\boldsymbol{B}}_{M,W}]_N$. Similar to the case for the DPSS's which can be obtained efficiently and numerically stably by computing the eigenvectors of a tridiagonal matrix [118], Grünbaum [57] showed that the prolate matrix $[\overline{\boldsymbol{B}}_{M,W}]_N$ commutes with a tridiagonal matrix, providing a stable and reliable method for computing the PDPSS's.

In the rest of this section, we assume $2K+1<M$, which is of practical interest for applications (e.g., [71, 57, 65]). Let $1\geq\lambda_N^{(0)}\geq\lambda_N^{(1)}\geq\cdots\lambda_N^{(N-1)}\geq 0$ denote the eigenvalues of $[\overline{\boldsymbol{B}}_{M,W}]_N$, where the upper and lower bounds follow because

$$\|\boldsymbol{x}\|^2\geq\boldsymbol{x}^*\overline{\boldsymbol{B}}_{M,W}\boldsymbol{x}=\sum_{k\in\mathcal{I}_K}|\widehat{\boldsymbol{x}}[k]|^2\geq 0$$

for all $\boldsymbol{x}\in\mathbb{C}^M$, indicating that the eigenvalues of $\overline{\boldsymbol{B}}_{M,W}$ are between 0 and 1 (and thus so are the eigenvalues of $[\overline{\boldsymbol{B}}_{M,W}]_N$ by the Sturmian separation theorem [67]). We note that when $2K+1>M$, it is possible that some eigenvalues $\lambda_N^{(\ell)}\geq 1$; see [137] for more discussion

on this.

We establish the following results (whose proof is given in Appendix A.7) concerning the eigenvalue distribution for $[\overline{\boldsymbol{B}}_{M,W}]_N$.

Theorem 3.2. *(Spectrum concentration for $[\overline{\boldsymbol{B}}_{M,W}]_N$) For any $M, N, K \in \mathbb{N}$, suppose $N < M$ and $W = \frac{2K+1}{2M} < \frac{1}{2}$. Then for any $\epsilon \in (0, \frac{1}{2})$, we have*

$$\lambda_N^{(2\lfloor NW \rfloor - R(N,M,\epsilon))} \geq 1-\epsilon, \;\; \lambda_N^{(2\lfloor NW \rfloor + R(N,M,\epsilon)+1)} \leq \epsilon,$$

and

$$\#\{\ell : \epsilon < \lambda_N^{(\ell)} < 1-\epsilon\} \leq 2R(N,M,\epsilon),$$

where

$$R(N,M,\epsilon) = \left(\frac{4}{\pi^2}\log(8N)+6\right)\log\left(\frac{16}{\epsilon}\right) + 2\max\left(\frac{-\log\left(\frac{\pi}{32}\left(\left(\frac{M}{N}\right)^2-1\right)\epsilon\right)}{\log\left(\frac{M}{N}\right)}, 0\right).$$

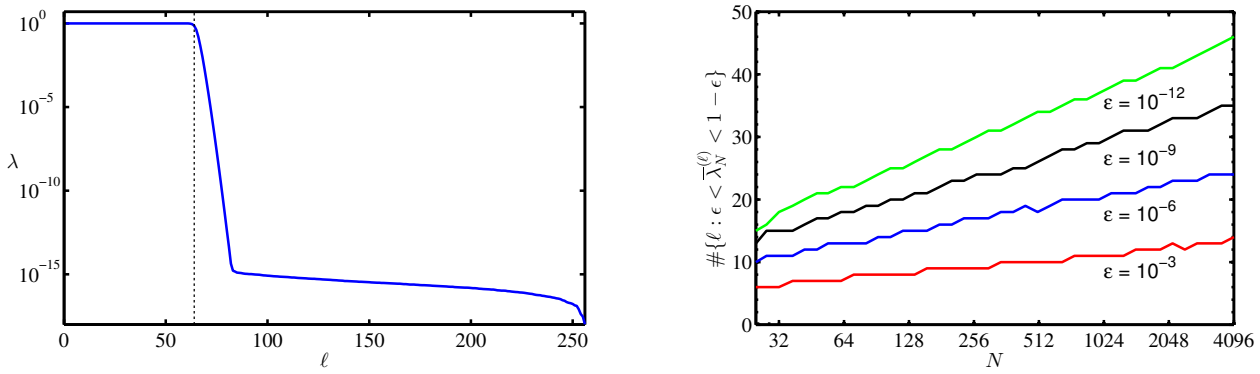
In words, Theorem 3.2 implies that the first $\approx \frac{(2K+1)N}{M}$ eigenvalues tend to cluster very close to one, while the remaining eigenvalues tend to concentrate about zero, after a narrow transition band of width $O(\log\frac{1}{\epsilon}\log N)$.[5] Figure 3.7(a) presents an example to illustrate this phenomenon. We note that this phenomenon has been utilized in [95] for efficiently computing Fourier extensions. A similar bound on the width of the transition band is given in [45], which shows

$$\#\left\{\ell : \epsilon \leq \lambda_N^{(\ell)} \leq 1-\epsilon\right\} \sim O(\log N). \tag{3.7}$$

This bound highlights the logarithmic dependence on N, but ignores the dependence on[6] ϵ. Theorem 3.2 improves (3.7) by showing also the logarithmic dependence on ϵ. Figure 3.7(b) shows the size of thetransition band for different N and ϵ, illustrating that the size is pro-

[5] $O(\cdot)$ denotes the standard "big-O" notation.

[6] With simple manipulation, this bound states $O(1/\epsilon)$ dependence on ϵ, which is quite large when ϵ is very small.



(a) (b)

Figure 3.7: (a) Eigenvalues of the prolate matrix $[\overline{\boldsymbol{B}}_{M,W}]_N$ with $M = 1024, N = 256, K = 128$ so that $\frac{(2K+1)N}{M} \approx 64$ (dashed line); (b) Width of the transition band $\#\{\ell : \epsilon < \lambda_N^{(\ell)} < 1-\epsilon\}$ for $N = \frac{1}{4}M, K = \frac{1}{8}M$, and $\epsilon = 10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}$.

portional to $\log \frac{1}{\epsilon} \log N$.

Theorem 3.2 also has implications regarding the distribution of singular values of submatrices of the DFT matrix. Specifically, let $\boldsymbol{F}_M$ be the normalized DFT matrix with entries given by

$$\boldsymbol{F}_M[m,n] = \frac{1}{\sqrt{M}} e^{-\jmath 2\pi \frac{mn}{M}}, \ m,n \in [M].$$

Let $L = \frac{M}{p}$ be an integer and let $\boldsymbol{F}_{M|p}$ denote an $L \times L$ submatrix of $\boldsymbol{F}_M$ obtained by deleting any consecutive $M-L$ columns and any consecutive $M-L$ rows of $\boldsymbol{F}_M$. Edelman et al. [45] proposed an approximate algorithm for DFT computations with lower communication cost based on the compressibility (low rank) of the blocks of $\boldsymbol{F}_M$, i.e., $\boldsymbol{F}_{M|p}$. Let $1 \geq \sigma^{(0)} \geq \cdots \geq \sigma^{(L-1)} \geq 0$ denote the singular values of $\boldsymbol{F}_{M|p}$. For any $\epsilon \in \left(0, \frac{1}{2}\right)$, similar to (3.7), $\#\left\{\ell : \sqrt{\epsilon} \leq \sigma^{(\ell)} \leq \sqrt{1-\epsilon}\right\} \sim O(\log L)$ is given in [45]. The following result (which is proved in Appendix A.8) establishes a finer non-asymptotic bound (that highlights the logarithmic dependence on ϵ) for this width.

Corollary 3.5. *For any $M, p \in \mathbb{N}$ such that $L = \frac{M}{p}$ is an integer, let $\boldsymbol{F}_{M|p}$ denote an $L \times L$ submatrix of the normalized DFT matrix $\boldsymbol{F}_M$ obtained by deleting any consecutive $M-L$ columns and any consecutive $M-L$ rows of $\boldsymbol{F}_M$. Let $\sigma^{(0)} \geq \cdots \geq \sigma^{(L-1)}$ denote the singular*

values of $\boldsymbol{F}_{M|p}$. *Then for any* $\epsilon \in (0, \frac{1}{2})$,

$$\sigma^{(2\lfloor \frac{L}{2p} \rfloor - R(L,M,\epsilon))} \geq \sqrt{1-\epsilon}, \quad \sigma^{(2\lfloor \frac{L}{2p} \rfloor + R(L,M,\epsilon)+1)} \leq \sqrt{\epsilon},$$

and

$$\# \left\{ \ell, \sqrt{\epsilon} < \sigma^{(\ell)} < \sqrt{1-\epsilon} \right\} \leq 2R\left(L, M, \epsilon\right),$$

where $R(\cdot,\cdot,\cdot)$ *is specified in Theorem 3.2.*

CHAPTER 4

ROAST: RAPID ORTHOGONAL APPROXIMATE SLEPIAN TRANSFORM

In the last chapter, we proposed a fast method for computing approximate projections onto the leading DPSS vectors and compressing a signal to the corresponding low dimension. Despite its favorable properties, the fast algorithm presented in Chapter 3 does not correspond to an orthogonal projection. In this chapter[7], we provide an alternative subspace—which enables a fast transform named Rapid Orthogonal Approximate Slepian Transform (ROAST)—for the discrete vector one obtains when collecting a finite set of uniform samples from a baseband analog signal (one of the parameterized subspace models listed in Section 1.2). The ROAST offers an orthogonal projection which is an approximation to the orthogonal projection onto the leading DPSS vectors. As such, the ROAST is guaranteed to accurately and compactly represent not only oversampled bandlimited signals but also the leading DPSS vectors themselves. Moreover, the subspace angle between the ROAST subspace and the corresponding DPSS subspace can be made arbitrarily small. The complexity of computing the representation of a signal using the ROAST is comparable to the FFT, which is much less than the complexity of using the DPSS basis vectors. We also give non-asymptotic results to guarantee that the proposed basis not only provides a very high degree of approximation accuracy in a mean-square error sense for bandlimited sample vectors, but also that it can provide high-quality approximations of all sampled sinusoids within the band of interest.

4.1 Construction of ROAST and Relation to the DPSS Subspace

In the last chapter, we demonstrated a fast method to approximately project an arbitrary vector onto the subspace spanned by the first slightly more than $2NW$ eigenvectors of $\boldsymbol{B}_{N,W}$

[7]This work is in collaboration with Mark A. Davenport, Santhosh Karnik, Justin Romberg, and Michael B. Wakin [144].

(i.e., the DPSS vectors) by utilizing the fact that the difference between $\boldsymbol{B}_{N,W}$ and $\boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ approximately has a rank of $O(\log N)$ (see Theorem 3.1). Note that, this approximate projection is not a true orthogonal projection onto any subspace. Here, we exhibit a subspace that captures most of the energy in the first $2NW$ DPSS vectors (and also the energy in sampled sinusoids within the band of interest), and this subspace has an orthogonal projector that can be applied efficiently to an arbitrary vector.

By utilizing the result that $\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ is approximately low rank and also that $\boldsymbol{F}_{N,W}$ can be applied to a vector efficiently with the FFT, we build a subspace with an orthonormal basis of the form

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \boldsymbol{Q}' \end{bmatrix},$$

where $\boldsymbol{Q}' \in \mathbb{C}^{N\times R}$ for some R that we can choose as desired. Let $\overline{\boldsymbol{F}}_{N,W}$ denote the $N \times (N - 2\lfloor NW \rfloor - 1)$ matrix with the highest frequency $N - 2\lfloor NW \rfloor - 1$ DFT vectors of length N. Thus $\boldsymbol{F}_N := \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W} \end{bmatrix}$ is the normalized DFT matrix. Since $\boldsymbol{Q}'$ must be orthogonal to $\boldsymbol{F}_{N,W}$ and the columns of $\boldsymbol{Q}'$ must be orthonormal, we can represent $\boldsymbol{Q}'$ by $\boldsymbol{Q}' = \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}$, where $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW \rfloor - 1)\times R}$ is orthonormal (one can verify that $\boldsymbol{F}_{N,W}^*\boldsymbol{Q}' = \boldsymbol{0}$ and $(\boldsymbol{Q}')^*\boldsymbol{Q}' = \mathbf{I}$). Thus, the desired orthogonal approximate Slepian basis is given as

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix}, \quad \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \mathbf{I}. \tag{4.1}$$

The optimal $\boldsymbol{V}$ is chosen such that the subspace spanned by $\boldsymbol{Q}$ captures the important DPSS vectors. (Since all the DPSS vectors $\boldsymbol{s}_{N,W}^{(0)}, \ldots, \boldsymbol{s}_{N,W}^{(N-1)}$ form an orthobasis for $\mathbb{C}^N$, no subspace of $\mathbb{C}^N$ can capture all of them except $\mathbb{C}^N$ itself.) To illustrate how we obtain $\boldsymbol{V}$, consider the following weighted least squares problem

$$\underset{\boldsymbol{Q}}{\text{minimize}} \quad \varrho(\boldsymbol{Q}) := \sum_{\ell=0}^{N-1} \lambda_{N,W}^{(\ell)} \left\| \boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{s}_{N,W}^{(\ell)} \right\|^2. \tag{4.2}$$

Here we use the DPSS eigenvalue $\lambda_{N,W}^{(\ell)}$ to weight the energy in the DPSS vector $\boldsymbol{s}_{N,W}^{(\ell)}$ that is not captured by $\boldsymbol{Q}$. The reason is that the larger the DPSS eigenvalue, the more concentration the corresponding DPSS vector has in the frequency domain, implying that the DPSS vector is more important in practical applications such as representing sampled bandlimited signals (see (2.20)). To solve (4.2), we rewrite $\varrho(\boldsymbol{Q})$ as

$$\begin{aligned}\varrho(\boldsymbol{Q}) &= \operatorname{trace}\left(\sum_{\ell=0}^{N-1}\lambda_{N,W}^{(\ell)}\boldsymbol{s}_{N,W}^{(\ell)}\left(\boldsymbol{s}_{N,W}^{(\ell)}\right)^{\mathrm{T}} - \boldsymbol{Q}\boldsymbol{Q}^{*}\lambda_{N,W}^{(\ell)}\boldsymbol{s}_{N,W}^{(\ell)}\left(\boldsymbol{s}_{N,W}^{(\ell)}\right)^{\mathrm{T}}\right)\\ &= \operatorname{trace}\left(\boldsymbol{B}_{N,W} - \boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{B}_{N,W}\right)\\ &= \int_{-W}^{W}\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2\,df,\end{aligned} \tag{4.3}$$

where the last line follows from (2.19). In other words, an orthonormal basis $\boldsymbol{Q}$ obtained by minimizing $\varrho(\boldsymbol{Q})$ is also an optimal basis to represent sampled bandlimited vectors in the MSE sense.

Plugging $\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}$ into the above equation yields

$$\varrho(\boldsymbol{Q}) = \operatorname{trace}\left(\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} - \boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}\right)$$

which suggests that setting $\boldsymbol{V}$ equal to the R dominant left singular vectors of $\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}$ (or $\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}$) results in a relatively small representation residual in the right hand of the above equation as long as $\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}$ has an effective rank of R. The following result provides a formal guarantee on this.

Theorem 4.1. *(Representation guarantee for DPSS vectors) Fix $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$. Let $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW\rfloor-1)\times R}$ be the R dominant left singular vectors of $\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}$. For any $\epsilon \in (0, \frac{1}{2})$, fix K to be such that $\lambda_{N,W}^{(K-1)} \geq \epsilon$. Then the orthobasis $\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}$ satisfies*

$$\begin{aligned}&\|\boldsymbol{S}_K\boldsymbol{S}_K^{*} - \boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{S}_K\boldsymbol{S}_K^{*}\|^2 \leq \epsilon,\\ &\|\boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{s}_{N,W}^{(\ell)}\|_2^2 \leq \epsilon,\end{aligned}$$

for all $l = 0, 1, \ldots, K-1$ *with* $R = \lceil C_N \log(15/\epsilon) \rceil$. *Here* C_N *is the constant specified in Theorem 3.1. By slightly increasing* R *to* $R = \lceil C_N \log(15N/\epsilon) \rceil$, *the subspace angle* $\Theta_{\boldsymbol{S}_K, \boldsymbol{Q}}$ *between the columns spaces of* $\boldsymbol{S}_K$ *and* $\boldsymbol{Q}$ *satisfies*

$$\cos(\Theta_{\boldsymbol{S}_K, \boldsymbol{Q}}) \geq \sqrt{1-\epsilon}.$$

The formal definition of (the largest principal) angle between two subspaces is given in Definition 2.3. In formally, if the subspace angle Θ is small, the two subspaces are nearly linearly dependent and one subspace is almost "contained" in the other subspace. Here, to guarantee that the column space of $\boldsymbol{S}_K$ is almost "contained" in the column space of $\boldsymbol{Q}$, one can make $\Theta_{\boldsymbol{S}_K, \boldsymbol{Q}}$ arbitrary small by increasing R. However, we note that we are not guaranteed that $\|\boldsymbol{Q}\boldsymbol{Q}^* - \boldsymbol{S}_K \boldsymbol{S}_K^*\|$ is small since in general $\|\boldsymbol{Q}\boldsymbol{Q}^* - \boldsymbol{S}_K \boldsymbol{S}_K^*\| = 1$ if $\boldsymbol{Q}$ and $\boldsymbol{S}_K$ have a different number of columns. Instead, we are guaranteed that the subspace spanned by the columns of $\boldsymbol{S}_K$ is approximately within the column space of $\boldsymbol{Q}$ and the subspace angle between the two subspaces is small by Theorem 4.1. We also note that the bound on $\|\boldsymbol{S}_K \boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{S}_K \boldsymbol{S}_K^*\|$ is useful since for any vector $\boldsymbol{a} \in \mathbb{C}^N$

$$\begin{aligned}
&\|\boldsymbol{a} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{a}\| \\
&\leq \|\boldsymbol{a} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{a}\| \\
&\leq \|\boldsymbol{a} - \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{a}\| + \|\boldsymbol{S}_K \boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{S}_K \boldsymbol{S}_K^*\| \|\boldsymbol{a}\| \\
&\leq \|\boldsymbol{a} - \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{a}\| + \sqrt{\epsilon} \|\boldsymbol{a}\|,
\end{aligned}$$

which implies any representation guarantee for $\boldsymbol{S}_K$ can be utilized for $\boldsymbol{Q}$.

4.2 Representations of Sampled Sinusoids and Oversampled Bandlimited Signals

As illustrated in (4.3), the orthonormal matrix obtained by minimizing $\varrho(\boldsymbol{Q})$ is also expected to accurately represent the sampled sinusoids within the band of interest in the MSE sense. This is formally established in the following results.

Theorem 4.2. *(Average representation error) Fix $W \in (0, \frac{1}{2})$ and $N \in \mathbb{N}$. Let $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW \rfloor - 1) \times R}$ contain the R dominant left singular vectors of $\overline{\boldsymbol{F}}_{N,W}^* \boldsymbol{B}_{N,W}$. Then for any $\epsilon \in (0, \frac{1}{2})$, the orthobasis $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W} \boldsymbol{V} \end{bmatrix}$ satisfies*

$$\int_{-W}^{W} \frac{\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} df \leq \epsilon$$

with

$$R = \max \left\{ \left\lceil C_N \log \left(\frac{15 C_N}{N \epsilon} \right) \right\rceil + 1, 0 \right\}.$$

Here C_N is the constant specified in Theorem 3.1.

A similar approximation guarantee holds for sampled vectors arising from sampling random bandlimited signals by using (2.21).

In [148] (see also Theorem 5.6 in Chapter 5), we rigorously show that every discrete-time sinusoid with a frequency $f \in [-W, W]$ is well-approximated by the DPSS basis $\boldsymbol{S}_K$ with K slightly larger than $2NW$. The proof is based on an asymptotic result on the DTFT of the DPSS basis functions (which are known as discrete prolate spheroidal wave functions (DPSWF's)) and the result is thus asymptotic. Here we use a different approach to obtain a non-asymptotic guarantee for approximating every discrete-time sinusoid with a frequency $f \in [-W, W]$. Noting that $\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2$ is differentiable everywhere, we first show that its derivative is bounded above by $2\pi N^2$. Then by utilizing the previous result on $\int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2 \, df$, one obtains a similar bound on $\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2$.

Theorem 4.3. *(Representation guarantee for pure sinusoids) Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. Also let $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW \rfloor - 1) \times R}$ be the R dominant left singular vectors of $\overline{\boldsymbol{F}}_{N,W}^* \boldsymbol{B}_{N,W}$. Then for any $\epsilon \in (0, \frac{1}{2})$, the orthobasis $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W} \boldsymbol{V} \end{bmatrix}$ satisfies*

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} \leq \epsilon$$

for all $f \in [-W, W]$ with

$$R = \max \left(\left\lceil C_N \log \left(\frac{60 \pi C_N}{\epsilon^2} \right) \right\rceil + 1, \ \left\lceil C_N \log \left(\frac{15 C_N}{NW \epsilon} \right) \right\rceil + 1 \right).$$

Here C_N is the constant specified in Theorem 3.1.

Remark 4.1. In [148] (see also Theorem 5.6 in Chapter 5), we show a similar but asymptotic result for the Slepian basis as follows. Fix $W \in (0, \frac{1}{2})$ and $\delta \in (0, \frac{1}{2W} - 1)$. Let $K = 2NW(1+\delta)$. Then there exist constants $\widetilde{C}_1, \widetilde{C}_2$ and $N_0 \in \mathbb{N}$ (which may depend on W and δ) such that

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} \leq \widetilde{C}_1 N^{3/2} e^{-\widetilde{C}_2 N}$$

for all $N \geq N_0$ and $f \in [-W, W]$. Compared with this result, Theorem 4.3 is non-asymptotic and provides a detail on the constants involved. However, we note that similar guarantees in Theorem 4.3 also holds for the DPSS basis by utilizing similar proof techniques for Theorem 4.3.

Theorem 4.4. *(Nonasymptotic representation guarantee for pure sinusoids with DPSS) Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. Also let $\boldsymbol{S}_K$ be an $N \times K$ matrix consists of the first K DPSS vector. Then for any $\epsilon \in (0, \frac{1}{2})$, the orthobasis $\boldsymbol{S}_K$ satisfies*

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{S}_K \boldsymbol{S}_K^* \boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} \leq \epsilon$$

for all $f \in [-W, W]$ with

$$K = 2NW + \max\left(\left\lceil C_N \log\left(\frac{60\pi C_N}{\epsilon^2}\right) \right\rceil + 1,\ \left\lceil C_N \log\left(\frac{15 C_N}{NW\epsilon}\right) \right\rceil + 1 \right).$$

Finally, we remark that for $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W} \boldsymbol{V} \end{bmatrix}$ with $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW \rfloor - 1) \times R}$, both $\boldsymbol{Q}$ and $\boldsymbol{Q}^*$ can be applied to a vector with computational complexity $O(N \log N + NR)$. As an example, for any $\boldsymbol{a} \in \mathbb{C}^N$, $\widetilde{\boldsymbol{a}} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W} \end{bmatrix}^* \boldsymbol{a}$ can be efficiently computed by the FFT with complexity $O(N \log N)$. Then $\boldsymbol{V}^* \widetilde{\boldsymbol{a}}_2$ can be computed via conventional matrix-vector multiplication with complexity $O(NR)$, where $\widetilde{\boldsymbol{a}}_2$ is the sub-vector obtained by taking the last $N - 2\lfloor NW \rfloor - 1$ entries of $\widetilde{\boldsymbol{a}}_2$. Thus the total computational complexity for $\boldsymbol{Q}^* \boldsymbol{a}$ is $O(N \log N + NR)$.

4.3 ROAST Construction with a Randomized Algorithm

We note that the DPSS vectors are not involved in constructing $\boldsymbol{V}$ and $\boldsymbol{Q}$. Directly computing $\boldsymbol{V}$ with Businger-Golub algorithm [14] requires $O(N(N-2\lfloor NW\rfloor-1)R)$ flops. Noting that $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}$ is effectively low rank, we can apply a fast randomized algorithm [60] to compute an approximate basis for the range of $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}$. Let $\boldsymbol{\Omega}$ be an $N\times P$ standard Gaussian matrix. We construct a matrix $\boldsymbol{V}'$ whose columns form an orthonormal basis for the range of $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}\boldsymbol{\Omega}$. By applying the FFT, the complexity of computing $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}\boldsymbol{\Omega}$ is $O(PN\log N)$. Computing an orthonormal basis for the range of $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}\boldsymbol{\Omega}$ requires $O(NP^2)$ flops. The following results establish the dimensionality of $\boldsymbol{V}'$ needed and the representation guarantee with the corresponding basis.

Theorem 4.5. *(Guarantee for randomized algorithm) Fix $N\in\mathbb{N}$ and $W\in(0,\frac{1}{2})$. Let $\boldsymbol{\Omega}$ be an $N\times P$ standard Gaussian matrix. Also let $\boldsymbol{V}$ be an orthonormal basis for the column space of the sample matrix $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}\boldsymbol{\Omega}$. For any $\epsilon\in(0,\frac{1}{2})$, fix K to be such that $\lambda^{(K-1)}_{N,W}\geq\epsilon$. Then the orthobasis $\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}$ has the following expression ability in expectation.*

- *Setting*

$$P=\left\lceil 2C_N\log\left(\frac{30+15e}{\epsilon}\right)\right\rceil+3,$$

we are guaranteed that

$$\begin{aligned}&\mathbb{E}\left[\left\|\boldsymbol{S}_K\boldsymbol{S}_K^*-\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\boldsymbol{S}_K^*\right\|^2\right]\leq\epsilon,\\&\mathbb{E}\left[\left\|\boldsymbol{s}^{(\ell)}_{N,W}-\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{s}^{(\ell)}_{N,W}\right\|_2^2\right]\leq\epsilon\end{aligned}$$

for all $l=0,1,\ldots,K-1$. By slightly increasing P to

$$P=\left\lceil 2C_N\log\left(\frac{(30+15e)\,N}{\epsilon}\right)\right\rceil+3,$$

we have

$$\mathbb{E}\left[\cos\left(\Theta_{\boldsymbol{S}_K,\boldsymbol{Q}}\right)\right] \geq \sqrt{1-N\epsilon}.$$

- *The sampled sinusoids within the band of interest in the least-squares sense can be captured well by* $\boldsymbol{Q}$ *in expectation:*

$$\mathbb{E}\left[\int_{-W}^{W} \frac{\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} df\right] \leq \epsilon$$

with

$$P = \left\lceil \frac{4}{3} C_N \log\left(\frac{15\sqrt{2C_N}}{\epsilon}\right) + \frac{7}{3} \right\rceil.$$

- *The orthonormal basis* $\boldsymbol{Q}$ *can also capture most of the energy in each pure sinusoid:*

$$\mathbb{E}\left[\frac{\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}\right] \leq \epsilon$$

for all $f \in [-W, W]$ *with*

$$P = \max\left(\left\lceil \frac{4}{3} C_N \log\left(\frac{60\pi N\sqrt{2C_N}}{\epsilon^2}\right) + \frac{7}{3} \right\rceil, \left\lceil \frac{4}{3} C_N \log\left(\frac{15\pi\sqrt{2C_N}}{W\epsilon}\right) + \frac{7}{3} \right\rceil\right).$$

Here $\mathbb{E}$ *denotes expectation with respect to the random matrix* $\boldsymbol{\Omega}$*.*

Remark 4.2. Using concentration of measure effects [60], we can argue that the results hold for a particular sampling matrix $\boldsymbol{\Omega}$ with high probability.

4.4 Benefits of an Orthonormal Basis

For any $\epsilon \in (0, \frac{1}{2})$, fix K to be such that $\lambda_{N,W}^{(K-1)} \geq \epsilon$. In the last chapter, we demonstrated a fast factorization of $\boldsymbol{S}_K\boldsymbol{S}_K^*$ by constructing two $N \times K'$ matrices $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ with $K' \leq \lceil 2NW \rceil + \left(\frac{12}{\pi^2}\log(8N) + 18\right)\log\left(\frac{15}{\epsilon}\right)$ such that

$$\|\boldsymbol{S}_K\boldsymbol{S}_K^* - \boldsymbol{T}_1\boldsymbol{T}_2^*\| \leq 2\epsilon.$$

Both $\boldsymbol{T}_2^*\boldsymbol{x}$ and $\boldsymbol{T}_1\boldsymbol{T}_2^*\boldsymbol{x}$ (as an approximation for $\boldsymbol{S}_K\boldsymbol{S}_K^*\boldsymbol{x}$) can be efficiently computed with $O(N \log N \log\frac{1}{\epsilon})$ operations.

However, neither $\boldsymbol{T}_1$ nor $\boldsymbol{T}_2$ is orthonormal and in general $\|\boldsymbol{T}_2^*\boldsymbol{x}\| \neq \|\boldsymbol{T}_1\boldsymbol{T}_2^*\boldsymbol{x}\|$ and $\|\boldsymbol{T}_2^*\boldsymbol{x}\| \neq \|\boldsymbol{T}_2\boldsymbol{T}_2^*\boldsymbol{x}\|$. Moreover, neither $\boldsymbol{T}_1$ nor $\boldsymbol{T}_2$ is well conditioned (i.e., both have a large condition number). In some applications, an approximate but orthonormal transform $\boldsymbol{Q}$ may be preferred, in order to ensure that $\|\boldsymbol{P}_{\boldsymbol{Q}}\boldsymbol{x}\| = \|\boldsymbol{Q}^*\boldsymbol{x}\|$ or that $\boldsymbol{Q}$ is well conditioned. We list two such stylized applications below.

Signal recovery

Suppose $\boldsymbol{x} \in \mathbb{C}^N$ is a sampled bandlimited signal with digital frequencies within the band $[-W, W]$ and we observe it through

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x},$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ $(2NW \leq M \leq N)$ is the sensing matrix. Knowing that $\boldsymbol{x}$ approximately lives in the subspace spanned by $\boldsymbol{S}_K$, we recover $\boldsymbol{x}$ by solving

$$\underset{\boldsymbol{\alpha}}{\text{minimize}}\, \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{S}_K\boldsymbol{\alpha}\|_F^2,$$

which is also a key part in compressive sensing of a discrete signal one obtains when collecting a finite set of uniform samples from a multiband analog signal [32]. The above least-squares problem is equivalent to the following system of linear equations

$$\boldsymbol{S}_K^*\boldsymbol{\Phi}^*\boldsymbol{\Phi}\boldsymbol{S}_K\boldsymbol{\alpha} = \boldsymbol{S}_K^*\boldsymbol{\Phi}^*\boldsymbol{y} \tag{4.4}$$

which can be solved by numerical algorithms such as conjugate gradient descent (CGD) [107]. The computational complexity of the CGD method depends on two factors: the convergence speed which depends on the condition number of the system $\boldsymbol{A} := \boldsymbol{S}_K^*\boldsymbol{\Phi}^*\boldsymbol{\Phi}\boldsymbol{S}_K$ and determines the number of iterations required, and the computational burden in each iteration mainly involving the application of $\boldsymbol{A}$ to a length-M vector. Utilizing a structured sensing matrix $\boldsymbol{\Phi}$ that has a fast implementation (such as the fast Johnson-Lindenstrauss transform [3]), we can efficiently implement $\boldsymbol{A}$ if we replace $\boldsymbol{S}_K$ by the fast transform $\boldsymbol{T}_1$ or $\boldsymbol{T}_2$ or the ROAST $\boldsymbol{Q}$ of the form (4.1). Unfortunately, both $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ have large condition number, resulting in slow convergence of the CGD method since the corresponding system $\boldsymbol{A}$ in general also

has large condition number. Thus, in this case, the orthonormal basis $\boldsymbol{Q}$ is preferable.

Line spectral estimation

Consider a measurement vector $\boldsymbol{y}$ consisting of a superposition of r sampled exponentials:

$$\boldsymbol{y} = \sum_{i=1}^{r} \alpha_i^\star \boldsymbol{e}_{f_i^\star}.$$

We may attempt to recover the frequencies $\{f_1^\star, \ldots, f_r^\star\}$ by solving the following nonlinear least squares problem

$$\{\widehat{f}_i, \widehat{\alpha}_i\} := \arg\min_{f_i, \alpha_i} \left\| \boldsymbol{y} - \sum_{i=1}^{r} \alpha_i \boldsymbol{e}_{f_i} \right\|^2. \tag{4.5}$$

Suppose we are given a priori knowledge that the frequencies $f_i^* \in [-W, W]$ for all $i \in \{1, \ldots, r\}$. Then we can reduce the computational cost of solving by (4.5) by projecting the measurements $\boldsymbol{y}$ onto the range space of $\boldsymbol{Q}$ [66]:

$$\{\overline{f}_i, \overline{\alpha}_i\} := \arg\min_{f_i, \alpha_i} \left\| \boldsymbol{P}_{\boldsymbol{Q}} \left(\boldsymbol{y} - \sum_{i=1}^{r} \alpha_i \boldsymbol{e}_{f_i} \right) \right\|^2 = \arg\min_{f_i, \alpha_i} \left\| \boldsymbol{Q}^* \left(\boldsymbol{y} - \sum_{i=1}^{r} \alpha_i \boldsymbol{e}_{f_i} \right) \right\|^2. \tag{4.6}$$

It is shown in [66] that the projected problem (4.6) has the same stationary points as the full problem (4.5) under certain conditions on the range space of $\boldsymbol{Q}$. When applying an optimization method like Gauss-Newton, the advantage of the projected problem (4.6) over the full problem (4.5) is that each optimization step is much cheaper since the projected Jacobian has much smaller size.

Based on this observation, for the general case where the frequencies lie in multiple bands, [66] provides an iterative algorithm that in each iteration, first finds one underlying band and projects the signal onto this band, then applies Gauss-Newton to solve the projected problem. We also note that our $\boldsymbol{Q}$ can be further reduce the computational cost in [66] since $\boldsymbol{Q}$ can be efficiently applied to a vector, while the orthonormal basis utilized in [66] is a numerical approximation (obtained by performing PCA on a set of sinusoids) to the Slepian basis $\boldsymbol{S}_K$.

4.5 Simulations

In this section, we present some experiments to illustrate the effectiveness of our proposed ROAST and ROAST-R (which is short for Rapid Orthogonal Approximate Slepian Transform with a Randomized algorithm for computing $\boldsymbol{V}$—see Section 4.3). Through this section, we use R (which is typically equal to $O(\log(N))$) to denote the the dimensionality of $\boldsymbol{V}$ for ROAST. For ROAST-R, we set P, the dimensionality of $\boldsymbol{V}$, as $P = R$ here.

For comparison, we also compute the projection onto the column space of $\boldsymbol{F}_{N,W+\frac{R}{2N}}$ which is the $N \times (2\lfloor NW \rfloor + 1 + R)$ DFT matrix with frequencies in $[-W - \frac{R}{2N}, W + \frac{R}{2N}]$. Such a projection is simply denoted by Sub-DFT. Note that the dimension of the column space of $\boldsymbol{F}_{N,W+\frac{R}{2N}}$ is $2\lfloor NW \rfloor + 1 + R$ and is equal to the dimension of the column space of $\boldsymbol{Q}$. In addition, the projection onto the column space of the leading DPSS vectors $\boldsymbol{S}_K$ is also computed and denoted by simply by DPSS in the legends of the figures. We also choose $K = 2\lfloor NW \rfloor + 1 + R$ so that all these subspaces have the same dimensionality.

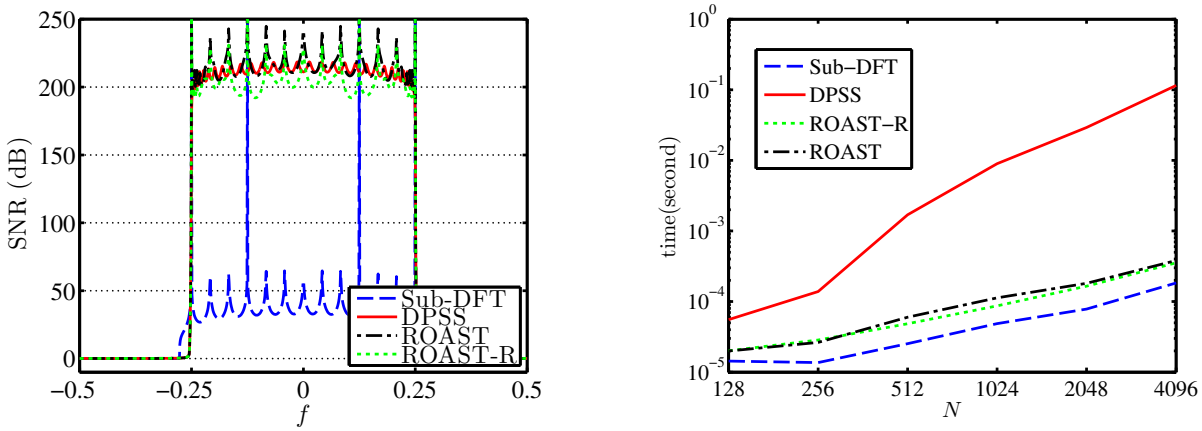
We quantify the ability of the different projections to capture a given signal $\boldsymbol{x} \in \mathbb{C}^N$ in terms of

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2} \right) \text{dB},$$

where $\widehat{\boldsymbol{x}}$ is the resulting projection of $\boldsymbol{x}$ by the above mentioned methods.

Figure 4.1(a) shows the SNR captured by different projections for various pure sinusoids $\boldsymbol{e}_f$. We observe that the DPSS basis, ROAST, ROAST-R and provide almost equal approximation performance for the pure sinusoids with frequencies in the band of interest. Also as guaranteed by Theorems 4.3, 4.4, and 4.5, any sinusoid in the band of interest can be well represented by the DPSS basis, ROAST and ROAST-R.

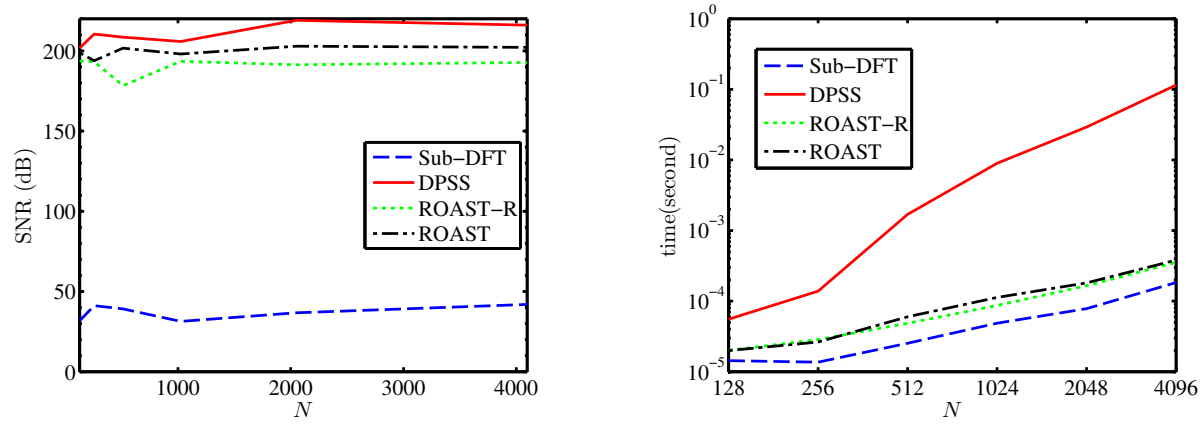
Also, we generate a sampled bandlimited signal $\boldsymbol{x}$ by adding 5000 complex exponentials with frequencies selected uniformly at random within the frequency band $[-W, W]$. Figure 4.1(b) shows the ability of the different projections to capture sampled bandlimited signals in terms of SNR. It can also be observed that the DPSS basis, ROAST and ROAST-R



(a) (b)

Figure 4.1: SNR captured by different projections (a) for pure sinusoids $\boldsymbol{e}_f$ with $R = 4\log(N)$ and (b) for a sampled bandlimited signal $\boldsymbol{x}$ with R ranging from 0 to $30 \approx 5\log(N)$. Here $N = 1024$, $W = \frac{1}{4}$.

provide almost equal approximation performance for sampled bandlimited signals.



(a) (b)

Figure 4.2: Comparison of different projections for a sampled bandlimited signal $\boldsymbol{x}$: (a) SNR as a function of N; (b) computation time as a function of N in a a logarithmic scale for both X-axis and Y-axis. In all plots, $W = \frac{1}{4}$ and $R = \lfloor 4\log(N) \rfloor$.

In addition, Figure 4.2 plots SNR as a function of dimension N and the relationship between the run time and N for the four projection methods. In this experiment, we fix $R = \lfloor 4\log(N) \rfloor$. As observed, the running time of DPSS has a quadratic increase, while

ROAST and ROAST-R[8] are nearly as fast as the DFT, but with much better approximation performance.

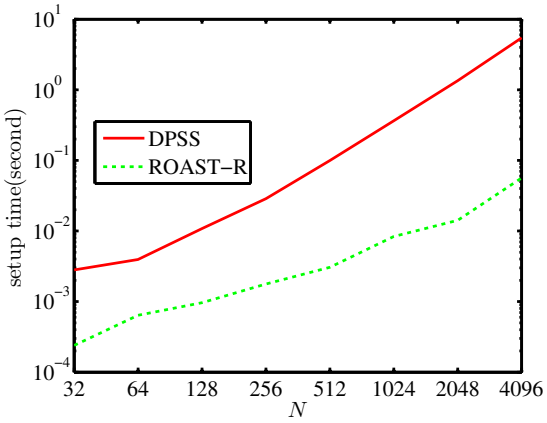


Figure 4.3: Precomputation time for DPSS and ROAST-R as a function of N in a logarithmic scale for both X-axis and Y-axis. In all plots, $W = \frac{1}{4}$ and $R = \lfloor 4 \log(N) \rfloor$.

Finally, we compare the precomputation time needed for DPSS basis and ROAST-R in Figure 4.3. For the DPSS basis, the first K DPSS vectors are precomputed with the Matlab command "dpss" (which actually computes the eigenvectors of a tridiagonal matrix with computational complexity of $O(N^2)$). For ROAST-R, the construction involves computing $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}\boldsymbol{\Omega}$ with $\boldsymbol{\Omega} \in \mathbb{R}^{N\times R}$ (which requires $O(RN \log N)$ operations) and computing an orthonormal basis for the range of $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}\boldsymbol{\Omega}$ (which requires $O(NR^2)$ operations). As can be seen in Figure 4.3, the precomputation time required by the DPSS grows roughly quadratically with N, while the precomputation time required by ROAST-R grows just faster than linearly in N.

[8]ROAST and ROAST-R are expected to have the same running time since these two transforms have the same dimensionality and form.

CHAPTER 5

APPROXIMATING SAMPLED SINUSOIDS AND MULTIBAND SIGNALS USING MULTIBAND MODULATED DPSS DICTIONARIES

In the last two chapters, we considered the parameterized subspace model where the signals of interested are obtained when collecting a finite set of uniform samples from baseband analog signals. In this chapter[9], we study possible dictionaries for representing the discrete vector one obtains when collecting a finite set of uniform samples from a multiband analog signal. To that end, we repeat (2.6) that

$$\mathbb{W} = [f_0 - W_0, f_0 + W_0] \cup [f_1 - W_1, f_1 + W_1] \cup \cdots \cup [f_{J-1} - W_{J-1}, f_{J-1} + W_{J-1}] \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]$$

is a union of J intervals. For each $i \in [J]$, define $\boldsymbol{\Psi}_i = [\boldsymbol{E}_{f_i} \boldsymbol{S}_{N,W_i}]_{k_i}$ for some value $k_i \in \{1, 2, \ldots, N\}$ that we can choose as desired. We construct the multiband modulated DPSS dictionary $\boldsymbol{\Psi}$ by concatenating these subdictionaries:

$$\boldsymbol{\Psi} := [\boldsymbol{\Psi}_0 \ \boldsymbol{\Psi}_1 \ \cdots \ \boldsymbol{\Psi}_{J-1}]. \tag{5.1}$$

We investigate the spectrum of the time- and multiband- limiting operator $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ (see Section 2.5 for the formal definition), as well as the efficiency of using $\boldsymbol{\Psi}$ to represent discrete-time sinusoids and sampled multiband signals. We also provide a stylized application in through-the-wall radar imaging at the end of this chapter.

5.1 Eigenvalues for Time- and Multiband-Limiting Operator

We begin by studying the eigenvalue concentration behavior of the operator $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ (and hence $\boldsymbol{B}_{N,\mathbb{W}}$), which reveals the effective dimensionality of the finite union of curves $\mathcal{M}_{\mathbb{W}} = \{\boldsymbol{e}_f\}_{f \in \mathbb{W}}$, where $\boldsymbol{e}_f$ is a length-N sinusoid defined in (2.1).

[9]This work is in collaboration with Michael B. Wakin [145, 148, 146, 147].

We first establish the following rough bound, which states that all the eigenvalues of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ are between 0 and 1. Its proof is given in Appendix C.1.

Lemma 5.1. *For any $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ and N, the operator $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ is positive-definite with eigenvalues*

$$1 > \lambda_{N,\mathbb{W}}^{(0)} \geq \lambda_{N,\mathbb{W}}^{(1)} \geq \cdots \geq \lambda_{N,\mathbb{W}}^{(N-1)} > 0$$

and

$$\sum_{l=0}^{N-1} \lambda_{N,\mathbb{W}}^{(\ell)} = N|\mathbb{W}|.$$

We denote the corresponding eigenvectors of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ by $\boldsymbol{u}_{N,\mathbb{W}}^{(0)}, \boldsymbol{u}_{N,\mathbb{W}}^{(1)}, \ldots, \boldsymbol{u}_{N,\mathbb{W}}^{(N-1)}$.

There is, in fact, a sharp transition in the distribution of the eigenvalues of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$. We establish this fact in the following theorem, which is proven in Appendix C.2.

Theorem 5.1. *Suppose $\mathbb{W}$ is a finite union of J pairwise disjoint intervals as defined in (2.6). For any $\varepsilon \in (0, \frac{1}{2})$, the number of eigenvalues of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ that are between ε and $1-\varepsilon$ satisfies*

$$\#\{\ell : \varepsilon \leq \lambda_{N,\mathbb{W}}^{(\ell)} \leq 1-\varepsilon\} \leq J \frac{\frac{2}{\pi^2}\log(N-1) + \frac{2}{\pi^2}\frac{2N-1}{N-1}}{\varepsilon(1-\varepsilon)}. \tag{5.2}$$

This result states that the number of eigenvalues in $[\epsilon, 1-\epsilon]$ is in the order of $\log(N)$ for any fixed $\epsilon \in (0, \frac{1}{2})$. Along with the following result which states that the number of eigenvalues of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ greater than $\frac{1}{2}$ equals $\approx N|\mathbb{W}|$, we conclude that the effective dimensionality of $\mathcal{M}_{\mathbb{W}}$ is approximately $N|\mathbb{W}| = \sum_i 2NW_i$. Its proof is given in Appendix C.3.

Theorem 5.2. *Let $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ be a finite union of J disjoint intervals having the form in (2.6). Denote by*

$$\iota_- = \#\{n \in \mathbb{Z} : -\lfloor \frac{N}{2} \rfloor \leq n \leq \lfloor \frac{N-1}{2} \rfloor,\ (\frac{n}{N} - \frac{1}{2N}, \frac{n}{N} + \frac{1}{2N}) \subset \mathbb{W}\}$$

and

$$\iota_+ = \#\{n \in \mathbb{Z} : -\lfloor \frac{N}{2} \rfloor \leq n \leq \lfloor \frac{N-1}{2} \rfloor,\ (\frac{n}{N} - \frac{1}{2N}, \frac{n}{N} + \frac{1}{2N}) \cap \mathbb{W} \neq \emptyset\}.$$

In particular, it holds that $\lfloor N|\mathbb{W}| \rfloor - 2J + 2 \leq \iota_- \leq \iota_+ \leq \lceil N|\mathbb{W}| \rceil + 2J - 2$. *Then the eigenvalues of the operator* $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ *satisfy*

$$\lambda_{N,\mathbb{W}}^{(\iota_- - 1)} \geq \frac{1}{2} \geq \lambda_{N,\mathbb{W}}^{(\iota_+)}.$$

Note that results similar to the above two theorems for time-frequency localization in the continuous domain have been established in [63, 70, 80]. Similar to the ideas used in [63], the key to proving Theorem 5.1 is to obtain an upper bound on the distance between the trace of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ and the sum of the squared eigenvalues of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$. Constructing an appropriate subspace with a carefully selected bandlimited sequence for the Weyl-Courant minimax characterization of eigenvalues is the key to proving Theorem 5.2. The proof techniques of [70, 80] form the basis of our analysis in Appendix C.3, but some modifications are required to extend their results to the discrete domain.

Similar to what happens in the single band case (when $J = 1$; see Lemma 2.1), the eigenvalues of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ have a distinctive behavior: the first $N|\mathbb{W}| = \sum_i 2NW_i$ eigenvalues tend to cluster very close to 1, while the remaining eigenvalues tend to cluster very close to 0, after a narrow transition. This is captured formally in the following result[10], whose proof is given in Appendix C.4.

Theorem 5.3. *Let* $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ *be a fixed finite union of* J *disjoint intervals having the form in (2.6).*

[10]Most of the results in this chapter build upon the asymptotic expressions for the DPSWF's [118] and hence are also asymptotic. It is possible to obtain nonasymptotic version of these results by utilizing the results in Chapters 3 and 4, especially Corollary 3.1 and Theorem 4.4.

1. *Fix $\epsilon \in (0,1)$. Then there exist constants $\overline{C}_1(\mathbb{W},\epsilon), \overline{C}_2(\mathbb{W},\epsilon)$ (which may depend on $\mathbb{W}$ and ϵ) and an integer $\overline{N}_0(\mathbb{W},\epsilon)$ (which may also depend on $\mathbb{W}$ and ϵ) such that*

$$\lambda_{N,\mathbb{W}}^{(\ell)} \geq 1 - \overline{C}_1(\mathbb{W},\epsilon) N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N}, \ \forall\ l \leq J - 1 + \sum_i \lfloor 2NW_i(1-\epsilon) \rfloor$$

for all $N \geq \overline{N}_0(\mathbb{W},\epsilon)$.

2. *Fix $\epsilon \in (0, \frac{1}{|\mathbb{W}|} - 1)$. Then there exist constants $\overline{C}_3(\mathbb{W},\epsilon), \overline{C}_4(\mathbb{W},\epsilon)$ (which may depend on $\mathbb{W}$ and ϵ) and an integer $\overline{N}_1(\mathbb{W},\epsilon)$ (which may also depend on $\mathbb{W}$ and ϵ) such that*

$$\lambda_{N,\mathbb{W}}^{(\ell)} \leq \overline{C}_3(\mathbb{W},\epsilon) e^{-\overline{C}_4(\mathbb{W},\epsilon)N}, \ \forall\ l \geq \sum_i \lceil 2NW_i(1+\epsilon) \rceil$$

for all $N \geq \overline{N}_1(\mathbb{W},\epsilon)$.

We point out that $\overline{N}_0(\mathbb{W},\epsilon) \geq \max\{N_0(W_i,\epsilon),\ \forall\ i \in [J]\}$, $\overline{C}_2(\mathbb{W},\epsilon) = \frac{\min\{C_2(W_i,\epsilon),\ \forall\ i \in [J]\}}{2}$, $\overline{C}_3(\mathbb{W},\epsilon) = J \max\{C_3(W_i,\epsilon),\ \forall\ i \in [J]\}$ and $\overline{C}_4(\mathbb{W},\epsilon) = \min\{C_4(W_i,\epsilon),\ \forall\ i \in [J]\}$, which will prove useful in our analysis below. Here $C_2(W_i,\epsilon)$, $C_3(W_i,\epsilon)$, and $C_4(W_i,\epsilon)$ are as specified in Lemma 2.1.

5.2 Multiband Modulated DPSS Dictionaries for Sampled Multiband Signals

Let $p \in \{1,2,\ldots,N\}$. Define

$$\boldsymbol{\Phi} := [\boldsymbol{u}_{N,\mathbb{W}}^{(0)}\ \boldsymbol{u}_{N,\mathbb{W}}^{(1)}\ \cdots\ \boldsymbol{u}_{N,\mathbb{W}}^{(p-1)}], \tag{5.3}$$

where $\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)},\ \forall\ l \in [N]$ are the eigenvectors of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$. Let $\boldsymbol{\Psi}$ be the multiband modulated DPSS dictionary defined in (5.1).

There are three main reasons why the dictionary $\boldsymbol{\Psi}$ may be useful representing sampled multiband signals. First, direct computation of $\boldsymbol{\Phi}$ is difficult due to the clustering of the eigenvalues of $\boldsymbol{B}_{N,\mathbb{W}}$. However, in the single band case, the matrix $\boldsymbol{B}_{N,W}$ is known to commute with a symmetric tridiagonal matrix that has well-separated eigenvalues, and hence its eigenvectors can be efficiently and stably computed [118]. Grünbaum [58] gave a certain

condition for a Toeplitz matrix to commute with a tridiagonal matrix with a simple spectrum. We can check that the matrix $\boldsymbol{B}_{N,\mathbb{W}}$ in general does not satisfy this condition, except for the case when $\mathbb{W}$ consists of only a single interval. However, we emphasize that $\boldsymbol{\Psi}$ is constructed simply by modulating DPSS's, which, again, can be computed efficiently.

Second, the multiband modulated DPSS dictionary $\boldsymbol{\Psi}$ provides an efficient representation for sampled multiband signals. Davenport and Wakin [32] provided theoretical guarantees into the use of this dictionary for sparsely representing sampled multiband signals and recovering sampled multiband signals from compressive measurements. We extend one of these guarantees in Section 5.2.3. Moreover, we confirm that a multiband modulated DPSS dictionary provides a high degree of approximation for all discrete-time sinusoids with frequencies in $\mathbb{W}$ in Section 5.2.2.

Third, as indicated by the results in Section 5.1, $\approx N|\mathbb{W}|$ dictionary atoms are necessary in order to achieve a high degree of approximation for the discrete-time sinusoids in a MSE sense. Our results, along with [32], show that the multiband modulated DPSS dictionary $\boldsymbol{\Psi}$ with $\approx N|\mathbb{W}|$ atoms can indeed approximate discrete-time sinusoids with high accuracy. In order to help explain this result, we first show that there is a near nesting relationship between the subspaces spanned by the columns of $\boldsymbol{\Psi}$ and by the columns of the optimal dictionary $\boldsymbol{\Phi}$.

5.2.1 The Subspace Angle

Our first guarantee considers the case where in constructing $\boldsymbol{\Psi}$, each k_i is chosen slightly smaller than $2NW_i$, and in constructing $\boldsymbol{\Phi}$, we take p to be slightly larger than $\sum_i 2NW_i$. In this case, we can guarantee that the subspace angle between $\mathcal{S}_{\boldsymbol{\Psi}}$ and $\mathcal{S}_{\boldsymbol{\Phi}}$ is small. The proof of the following result in given in Appendix C.5.

Theorem 5.4. *Let $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ be a fixed finite union of J disjoint intervals having the form in (2.6). Fix $\epsilon \in (0, \min\{1, \frac{1}{|\mathbb{W}|} - 1\})$. Let $p = \sum_i \lceil 2NW_i(1+\epsilon) \rceil$ and $\boldsymbol{\Phi}$ be the $N \times p$ matrix defined in (5.3). Also let $k_i \leq \lfloor 2NW_i(1-\epsilon) \rfloor, \forall i \in [J]$ and $\boldsymbol{\Psi}$ be the matrix defined*

in (5.1). Then for any column $\boldsymbol{\psi}$ in $\boldsymbol{\Psi}$,

$$||\boldsymbol{\psi} - \boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{\psi}||_2^2 \leq \frac{2\widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N}}{\left(1 - \widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N} - \overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}\right)^2} =: \kappa_1(N,\mathbb{W},\epsilon)$$

and

$$\cos(\Theta_{\mathcal{S}_{\boldsymbol{\Psi}}\mathcal{S}_{\boldsymbol{\Phi}}}) \geq \sqrt{\frac{1 - \kappa_1(N,\mathbb{W},\epsilon) - N\sqrt{\kappa_1(N,\mathbb{W},\epsilon)} - 3N\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)}e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}}{1 + 3N\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)}e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}}} \tag{5.4}$$

if $N \geq \max\{\overline{N}_0(\mathbb{W},\epsilon), \overline{N}_1(\mathbb{W},\epsilon)\}$. Here $\widetilde{C}_1(\mathbb{W},\epsilon) = \max\{C_1(W_i,\epsilon),\ \forall\ i \in [J]\}$, $\widetilde{C}_2(\mathbb{W},\epsilon) = \min\{C_2(W_i,\epsilon),\ \forall\ i \in [J]\}$, $\overline{N}_0(\mathbb{W},\epsilon)$, $\overline{N}_1(\mathbb{W},\epsilon)$, $\overline{C}_3(\mathbb{W},\epsilon)$, and $\overline{C}_4(\mathbb{W},\epsilon)$ are the constants specified in Theorem 5.3, and $C_1(W_i,\epsilon)$ and $C_2(W_i,\epsilon)$ are the constants specified in Lemma 2.1.

We can also guarantee that the subspace angle between $\mathcal{S}_{\boldsymbol{\Psi}}$ and $\mathcal{S}_{\boldsymbol{\Phi}}$ is small if, in constructing $\boldsymbol{\Psi}$, each k_i is chosen slightly larger than $2NW_i$, and in constructing $\boldsymbol{\Phi}$, we take p to be slightly smaller than $\sum_i 2NW_i$. This result is established in Corollary 5.1 (whose proof is given in Appendix C.7), which follows from Theorem 5.5 (whose proof is given in Appendix C.6).

Theorem 5.5. *Let $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ be a finite union of J disjoint intervals having the form in (2.6). Given some values $k_i \in \{1, 2, \ldots, N\}, \forall i \in [J]$, let $\boldsymbol{\Psi}$ be the matrix defined in (5.1). Then*

$$||\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}||_2 \geq \lambda_{N,\mathbb{W}}^{(\ell)} - \sum_{i=0}^{J-1}\sum_{l_i=k_i}^{N-1}\lambda_{N,W_i}^{(l_i)}$$

for all $l \in \{0, 1, \ldots, N-1\}$.

Corollary 5.1. *Let $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ be a fixed finite union of J disjoint intervals having the form in (2.6). Fix $\epsilon \in (0, \min\{1, \frac{1}{|\mathbb{W}|} - 1\})$. Let $p \leq J - 1 + \sum_i \lfloor 2NW_i(1-\epsilon) \rfloor$ and $\boldsymbol{\Phi}$ be the $N \times p$ matrix defined in (5.3). Also let $k_i = \lceil 2NW_i(1+\epsilon) \rceil, \forall i \in [J]$ and $\boldsymbol{\Psi}$ be the matrix*

defined in (5.1). Then for any column $\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}$ *in* $\boldsymbol{\Phi}$*,*

$$||\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}||_2 \geq 1 - \overline{C}_1(\mathbb{W},\epsilon)N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N} - N\overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}$$

and

$$\cos(\Theta_{\mathcal{S}_{\boldsymbol{\Psi}}\mathcal{S}_{\boldsymbol{\Phi}}}) \geq \sqrt{1 - 2\kappa_2(N,\mathbb{W},\epsilon) + \kappa_2^2(N,\mathbb{W},\epsilon) - N\sqrt{2\kappa_2(N,\mathbb{W},\epsilon) - \kappa_2^2(N,\mathbb{W},\epsilon)}} \quad (5.5)$$

for all $N \geq \max\{\overline{N}_0(\mathbb{W},\epsilon), \overline{N}_1(\mathbb{W},\epsilon)\}$*, where* $\overline{N}_i(\mathbb{W},\epsilon)$ *and* $\overline{C}_i(\mathbb{W},\epsilon)$ *are constants specified in Theorem 5.3, and* $\kappa_2(N,\mathbb{W},\epsilon)$ *is defined as* $\kappa_2(N,\mathbb{W},\epsilon) := \overline{C}_1(\mathbb{W},\epsilon)N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N} + N\overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}$.

Although our results hold for scenarios where one dictionary contains $\sum_i \lfloor 2NW_i(1-\epsilon)\rfloor$ atoms while another one has $\sum_i \lceil 2NW_i(1+\epsilon)\rceil$ atoms, we note that these dimensions can be made very close by choosing ϵ sufficiently small.[11]

5.2.2 Approximation Quality for Discrete-time Sinusoids

The above results show that $\boldsymbol{\Psi}$ spans nearly the same space as $\boldsymbol{\Phi}$ in the case where both dictionaries contain $\approx N|\mathbb{W}|$ columns. In this section, we investigate the approximation quality of $\boldsymbol{\Psi}$ for discrete-time sinusoids with frequencies in the bands of interest. Then, in the next section, we investigate the approximation quality of $\boldsymbol{\Psi}$ for sampled multiband signals.

We first prove that a single band dictionary with slightly more than $2NW$ baseband DPSS vectors can capture almost all of the energy in any sinusoid with a frequency in $[-W, W]$. Our analysis is based upon an expression for the DTFT of the DPSS vectors proposed in [118]. We review this result in Appendix C.8.

[11] Though a small ϵ may require N large enough such that our results hold, $\frac{\sum_i \lfloor 2NW_i(1-\epsilon)\rfloor}{\sum_i \lceil 2NW_i(1+\epsilon)\rceil}$ (the ratio between the sizes of the two dictionaries) may become close to 1.

Theorem 5.6. *Fix $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2W} - 1)$. Let $W' = \frac{1}{2} - W$, $\epsilon' = \frac{W}{\frac{1}{2}-W}\epsilon$ and $k = 2NW(1+\epsilon)$. Then there exists a constant $C_9(W', \epsilon')$ (which may depend on W' and ϵ') such that*

$$||\boldsymbol{e}_f - \boldsymbol{P}_{[\boldsymbol{S}_{N,W}]_k}\boldsymbol{e}_f||_2^2 \leq C_9(W', \epsilon')N^{5/2}e^{-C_2(W',\epsilon')N}, \ \forall |f| \leq W$$

for all $N \geq N_0(W', \epsilon')$, where $N_0(W', \epsilon')$ and $C_2(W', \epsilon')$ are constants defined in Lemma 2.1.

The proof is given in Appendix C.9. Similar to Theorem 4.4, Theorem 5.6 rigorously shows that asymptotically *every* discrete-time sinusoid with a frequency $f \in [-W, W]$ is well-approximated by a DPSS basis $[\boldsymbol{S}_{N,W}]_k$ with k slightly larger than $2NW$. This result extends the approximation guarantee in a MSE sense presented in [32]. We now extend this result for the multiband modulated DPSS dictionary. The proof of the following result is given in Appendix C.10

Corollary 5.2. *Let $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$ be a fixed finite union of J disjoint intervals having the form in (2.6). Fix $\epsilon \in (0, \frac{1}{|\mathbb{W}|} - 1)$. Let $k_i = 2NW_i(1+\epsilon), \forall i \in [J]$ and $\boldsymbol{\Psi}$ be the matrix defined in (5.1). Then there exist constants $C_{10}(\mathbb{W}, \epsilon)$ and $C_{11}(\mathbb{W}, \epsilon)$ (which may depend on $\mathbb{W}$ and ϵ) and an integer $N_2(\mathbb{W}, \epsilon)$ (which may also depend on $\mathbb{W}$ and ϵ) such that*

$$||\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f||_2^2 \leq C_{10}(\mathbb{W}, \epsilon)N^{5/2}e^{-C_{11}(\mathbb{W},\epsilon)N}, \ \forall f \in \mathbb{W} \tag{5.6}$$

for all $N \geq N_2(\mathbb{W}, \epsilon)$.

5.2.3 Approximation Quality for Sampled Multiband Signals (Statistical Analysis)

As indicated in [32], in a probabilistic sense, most finite-length sample vectors arising from multiband analog signals can be well-approximated by the multiband modulated DPSS dictionary. In this final section, we generalize the result [32, Theorem 4.4] to sampled multiband signals where each band has a possibly different width.

Theorem 5.7. *Suppose for each $i \in [J]$, $x_i(t)$ is a continuous-time, zero-mean, wide sense stationary random process with power spectrum*

$$P_{x_i}(F) = \begin{cases} \frac{1}{\sum_{i=0}^{J-1} B_{\text{band}_i}}, & F_i - \frac{B_{\text{band}_i}}{2} \leq F \leq F_i + \frac{B_{\text{band}_i}}{2} \\ 0, & \text{otherwise,} \end{cases}, \tag{5.7}$$

and furthermore suppose $x_0(t), x_1(t), \ldots, x_{J-1}(t)$ are independent and jointly wide sense stationary. Let T_s denote a sampling interval chosen to satisfy the minimum Nyquist sampling rate, which means $T_s \leq \frac{1}{B_{\text{nyq}}} := 1/\left(2\max\left\{\left|F_i \pm \frac{B_{\text{band}_i}}{2}\right|, \ \forall\ i \in [J]\right\}\right)$. Let $\boldsymbol{x}_i = [x_i(0)\ x_i(T_s)\ \ldots\ x_i((N-1)T_s)]^T \in \mathbb{C}^N$ denote a finite vector of samples acquired from $x_i(t)$ and let $\boldsymbol{x} = \sum_{i=1}^{J} \boldsymbol{x}_i$. Set $f_i = F_i T_s$ and $W_i = \frac{B_{\text{band}_i} T_s}{2}$. Let $\boldsymbol{\Psi}$ be the matrix defined in (5.1) for some given k_i. Then

$$\mathbb{E}[\|\boldsymbol{x} - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}\|_2^2] \leq \frac{1}{|\mathbb{W}|} \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \lambda_{N,W_i}^{(\ell_i)}, \tag{5.8}$$

where $\mathbb{E}[\|\boldsymbol{x}\|_2^2] = N$.

The proof is given in Appendix C.11. The right hand side of (5.8) can be made small by choosing $k_i \approx 2NW_i$ for each $i \in [J]$; recall Lemma 2.1. Aside from allowing for different band widths, the above result improves the upper bound of [32, Theorem 4.4] by a factor of J.

Finally, the following result establishes a deterministic guarantee for the approximation of sampled multiband signals using a multiband modulated DPSS dictionary with $\approx N|\mathbb{W}|$ atoms.

Corollary 5.3. *Suppose x is a continuous-time signal with Fourier transform $X(F)$ supported on $\mathbb{F} = \bigcup_{i=0}^{J-1} [F_i - B_{band_i}/2, F_i + B_{band_i}/2]$, i.e.,*

$$x(t) = \int_{\mathbb{F}} X(F) e^{j2\pi F t} dF.$$

Let $\boldsymbol{x} = [x(0)\ x(T_s)\ \ldots\ x((N-1)T_s)]^T \in \mathbb{C}^N$ denote a finite vector of samples acquired from $x(t)$ with a sampling interval of $T_s \leq 1/(2\max\{|F_c \pm \frac{B_{\text{band}}}{2}|\})$. Let $W_i = T_s B_{band_i}/2$,

$f_i = T_s F_i$ for all $i \in [J]$, and $\mathbb{W} = \bigcup_{i=0}^{J-1} [f_i - W_i, f_i + W_i]$. Fix $\epsilon \in (0, \frac{1}{|\mathbb{W}|} - 1)$. Let $k_i = 2NW_i(1+\epsilon), \forall i \in [J]$ and let $\boldsymbol{\Psi}$ be the matrix defined in (5.1). Then

$$||\boldsymbol{x} - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}||_2^2 \leq \left(\int_{\mathbb{W}} |\widetilde{x}(f)|^2 \, df \right) \cdot C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W},\epsilon)N} \tag{5.9}$$

for all $N \geq N_2(\mathbb{W}, \epsilon)$, where $N_2(\mathbb{W}, \epsilon)$, $C_{10}(\mathbb{W}, \epsilon)$ and $C_{11}(\mathbb{W}, \epsilon)$ are constants specified in Corollary 5.2.

The proof is given in Appendix C.12. Corollary 5.3 can be applied in various settings:

- The sequence $x[n]$ encountered in most practical problems has finite energy. For example, if we assume that $\int_{\mathbb{W}} |\widetilde{x}(f)|^2 df \leq 1$, we conclude that $||\boldsymbol{x} - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}||_2^2 \leq C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W},\epsilon)N}$.

- Moreover, in some practical problems, the finite-energy sequence $x[n]$ may be approximately time-limited to the index range $n = 0, 1, \ldots, N-1$ such that for some δ, $||\boldsymbol{x}||_2^2 = ||\mathcal{I}_N(x)||_2^2 \geq (1-\delta)||x||_2^2$. In this case, (5.9) guarantees that

$$\begin{aligned} \frac{||\boldsymbol{x} - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}||_2^2}{||\boldsymbol{x}||_2^2} &\leq \frac{\int_{\mathbb{W}} |\widetilde{x}(f)|^2 df}{||\boldsymbol{x}||_2^2} \cdot C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W},\epsilon)N} \\ &\leq \frac{1}{1-\delta} C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W},\epsilon)N}, \end{aligned} \tag{5.10}$$

  where the last inequality follows from Parseval's theorem that $||x||_2^2 = \int_{\mathbb{W}} |\widetilde{x}(f)|^2 df$.

Along with the result proved in [32] that samples from a time-limited sequence which is approximately bandlimited to the bands of interest can be well-approximated by the multiband modulated DPSS dictionary, we conclude that the multiband modulated DPSS dictionary is useful for most practical problems involving representing sampled multiband signals.

However, we point out that not *all* sampled multiband signals can be well-approximated by the multiband modulated DPSS dictionary. To illustrate this, consider the simple case where $\mathbb{W}$ reduces to a single band $[-W, W]$. Recalling that the infinite-length DPSS's are

strictly bandlimited, it follows that each of the DPSS vectors can be obtained by sampling and time-limiting some strictly bandlimited analog signal. Nevertheless, for all $l \geq k$, we will have

$$\frac{||\boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{P}_{[\boldsymbol{S}_{N,W}]_k} \boldsymbol{s}_{N,W}^{(\ell)}||_2}{||\boldsymbol{s}_{N,W}^{(\ell)}||_2} = 1 \tag{5.11}$$

even when we choose $k = 2NW(1+\epsilon)$. In this case, the approximation guarantee in (5.11) is much worse than what appears in (5.10). Such examples are pathological, however: the infinite sequence $s_{N,W}^{(\ell)}$ has energy $||s_{N,W}^{(\ell)}||_2^2 = (\lambda_{N,W}^{(\ell)})^{-1}$, which according to Lemma 2.1 is exponentially large when $l \geq 2NW(1+\epsilon)$, and yet the energy of the sampled vector $||\boldsymbol{s}_{N,W}^{(\ell)}||_2^2$ is only 1. Moreover, the spectrum of the infinite sequence $s_{N,W}^{(\ell)}$ is entirely concentrated in the band $[-W, W]$ while the spectrum of the time-limited sequence $\mathcal{T}_N(s_{N,W}^{(\ell)})$ is almost entirely contained outside the band $[-W, W]$, and so on. We refer the reader to [32] for additional discussion of this topic.

5.3 Stylized Application: Through-the-Wall Radar Imaging

As an particular application, in this section, we investigate the wall and target return subspaces both (i) for each antenna element separately and (ii) jointly for all antenna elements in through-the-wall radar imaging. We mainly utilize Theorem 5.6 that all sampled sinusoids in the targeted band can be represented well by the dictionary $\boldsymbol{Q}$. To simply the notations (especially the constants) involved in Theorem 5.6, we restate it as follows.

Theorem 5.8. *Fix $W \in (0, \frac{1}{2})$ and $f_c \in \mathbb{R}$. Define $\boldsymbol{Q} := [\boldsymbol{E}_{f_c} \boldsymbol{S}_{N,W}]_J$. For fixed $\epsilon \in (0,1)$, choose $J = 2NW(1+\epsilon)$. Then there exist constants C_1, C_2 (where C_1, C_2 may depend on W and ϵ) such that for all $N \geq N_0$*

$$||\boldsymbol{P}_{\boldsymbol{Q}} \boldsymbol{e}_f||_2 \leq C_1 N^{5/4} e^{-C_2 N}, \ \forall \ f \in [f_c - W, f_c + W].$$

In a nutshell, this result states that (i) the effective dimensionality of the subspace spanned by $\{\boldsymbol{e}_f\}_{f \in [f_c - W, f_c + W]}$ is $2NW$, and (ii) the modulated DPSS vectors provide a

basis for this subspace.

5.3.1 Problem Setup

We consider an M-element synthetic linear aperture that transmits waveforms and receives the reflected signals. We assume that each transceiver transmits and receives a stepped-frequency signal consisting of N frequencies equispaced over the band $[f_0, f_{N-1}]$ with the frequency step size $\Delta F := \frac{f_{N-1}-f_0}{N-1}$, i.e., $f_n = f_0 + n\Delta F$. Further, we assume monostatic operation in which the transmitter and receiver are collocated as viewed from the target (i.e., the same antenna is used to transmit and receive) and after the antenna obtains the measurements in one location, we move it to the next location. To simply the notation, we suppose the antennas are parallel to the wall. According to [2], we can model the wall return at the m-th antenna and the n-th frequency as

$$\boldsymbol{r}_m^w[n] := \sum_{l=0}^{L} \vartheta_l e^{-j2\pi f_n t_l}, \ \forall\ m \in [M],\ n \in [N]. \tag{5.12}$$

Here, $[N]$ denotes the set $\{0, 1, \ldots, N-1\}$ for any natural number $N \in \mathbb{N}$; ϑ_0 is the complex reflectivity of the wall; $\vartheta_l,\ l \geq 1$ represents the complex reflectivity corresponding to the l-th wall reverberation and decreases with l; L denotes the number of wall reverberations; t_0 is the direct two-way travel time between the wall and the antenna; and $t_l,\ l \geq 1$ is the delay associated with the l-th wall return to the antenna.

Suppose there are K targets behind the wall. The target return observed by the m-th antenna at the n-th frequency can be expressed as

$$\boldsymbol{r}_m^t[n] := \sum_{k=1}^{K} \boldsymbol{r}_{k,m}^t[n], \tag{5.13}$$

where $\boldsymbol{r}_{k,m}^t[n] := \int_{\tau_{k,m}^{\min}}^{\tau_{k,m}^{\max}} \sigma_k(\tau) e^{-j2\pi f_n \tau} d\tau$. Here, $\sigma_k(\tau)$ is the complex reflectivity function of the k-th target (we assume the target reflectivity is independent of frequency), and $\tau_{k,m}^{\min}$ and $\tau_{k,m}^{\max}$ are the minimum and maximum two-way travel times between the k-th target

and the m-th antenna, respectively. Note that the return from point targets degenerates to $\boldsymbol{r}_m^t[n] = \sum_{k=1}^K \sigma_k e^{-j2\pi f_n \tau_{k,m}}$, where $\tau_{k,m}$ is the two-way travel time between the k-th point target and the m-th antenna [2].

The measurement $\boldsymbol{y}_m := \boldsymbol{r}_m^w + \boldsymbol{r}_m^t$ received by m-th antenna consists both wall and target return. Define $\boldsymbol{r}^w = [(\boldsymbol{r}_0^w)^H \ \cdots \ (\boldsymbol{r}_{M-1}^w)^H]^H$ and $\boldsymbol{r}^t = [(\boldsymbol{r}_0^t)^H \ \cdots \ (\boldsymbol{r}_{M-1}^t)^H]^H$. Here H represents the conjugate transpose. The measurements $\{\boldsymbol{y}_m\}_{m\in[M]}$ are arranged into an $MN \times 1$ vector $\boldsymbol{y} = \boldsymbol{r}^w + \boldsymbol{r}^t$. From the measurements $\boldsymbol{y}$, our goal is to detect or localize the potential targets.

5.3.2 Wall Return Subspace

The dimensionality of the wall return subspace

If we consider only the direct wall return, the wall return defined in (5.12) reduces to

$$\boldsymbol{r}_m^w[n] = \vartheta_0 e^{-j2\pi f_n t_0} = \vartheta_0 e^{-j2\pi f_0 t_0} e^{-j2\pi n \Delta F t_0}.$$

In this simple case, the wall return $\boldsymbol{r}_m^w$ lives in a 1-dimensional subspace spanned by the basis vector $\boldsymbol{e}_{-\Delta F t_0}$.

More generally, the wall return in (5.12) can be rewritten as

$$\boldsymbol{r}_m^w[n] = \sum_{l=0}^{L} \vartheta_l e^{-j2\pi f_0 t_l} e^{-j2\pi n t_l \Delta F}. \tag{5.14}$$

From Theorem 5.8, we expect that the wall return $\boldsymbol{r}_m^w$ at one antenna will approximately live within a low-dimensional subspace because (5.14) indicates that $\boldsymbol{r}_m^w$ can be viewed as a linear combination of sampled exponentials $\boldsymbol{e}_f$ with $f \in [-t_L \Delta F, -t_0 \Delta F]$. Accordingly, define the dictionary of modulated DPSS vectors

$$\boldsymbol{D}_m := [\boldsymbol{E}_{-\Delta F(t_L+t_0)/2} \boldsymbol{S}_{N,\Delta F(t_L-t_0)/2}]_{J^w}$$

for some value of $J^w \in \{1, 2, \ldots, N\}$.

Corollary 5.4. *Fix $\epsilon \in (0,1)$. Choose $J^w = N(t_L - t_0)\Delta F(1+\epsilon)$. Then there exist constants C_1, C_2 and an integer N_0 such that for all $N \geq N_0$*

$$\|\boldsymbol{P}_{\boldsymbol{D}_m}\boldsymbol{r}_m^w\|_2 \leq \sum_{l=0}^{L} \vartheta_l C_1 N^{5/4} e^{-C_2 N}.$$

The proof follows directly from Theorem 5.8. The above result indicates that $\boldsymbol{r}_m^w$ is approximately within $\mathcal{S}_{\boldsymbol{D}_m}$, the column space of $\boldsymbol{D}_m$ when we set $J^w = N(t_L - t_0)\Delta F(1+\epsilon)$. Define the $MN \times MJ^w$ block diagonal matrix $\boldsymbol{D} := \text{diag}(\boldsymbol{D}_0, \ldots, \boldsymbol{D}_{M-1})$.

Corollary 5.5. *Fix $\epsilon \in (0,1)$. Choose $J^w = N(t_L - t_0)\Delta F(1+\epsilon)$. Then there exist constants C_1, C_2 and an integer N_0 such that for all $N \geq N_0$*

$$\|\boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^w\|_2 \leq M \sum_{l=0}^{L} \vartheta_l C_1 N^{5/4} e^{-C_2 N}.$$

In words, the complete wall return $\boldsymbol{r}^w$ lives approximately within $\mathcal{S}_{\boldsymbol{D}}$, the subspace spanned by the columns of $\boldsymbol{D}$. The dimension of $\mathcal{S}_{\boldsymbol{D}}$ is MJ^w.

Before moving on, we note that the electrical properties of the wall material, which directly determine $\{t_l\}_{l=1}^L$, may not be known in advance. The dictionary $\boldsymbol{D}_m$ cannot capture the wall return completely if t_L is chosen too small. On the other hand, choosing t_L too large may result in a dictionary $\boldsymbol{D}_m$ that also captures some energy from target returns behind the wall. Since simulations have indicated that almost all walls have dominant reverberations up to 1.5m behind the wall [2], we use the same strategy as [2] in that we mitigate the wall reverberations up to 1.5m behind the wall. Note that we can still detect some targets located less than 1.5m behind the wall as long as there exist some antennas that have distance larger than 1.5m from these targets.

Because we assume the antennas are parallel to the wall, the wall return $\boldsymbol{r}_m^w$ is identical for different m. Therefore, the wall return $\boldsymbol{r}^w$ actually lives within a subspace which has much smaller dimension than $\mathcal{S}_{\boldsymbol{D}}$. Define $\widehat{\boldsymbol{D}} := \frac{1}{\sqrt{M}}[\boldsymbol{D}_0^H \ \boldsymbol{D}_1^H \ \cdots \ \boldsymbol{D}_{M-1}^H]^H$.

Corollary 5.6. *Fix $\epsilon \in (0,1)$. Choose $J^w = N(t_L - t_0)\Delta F(1+\epsilon)$. Then there exist constants C_1, C_2 and an integer N_0 such that for all $N \geq N_0$*

$$\|\boldsymbol{P}_{\widehat{\boldsymbol{D}}}\boldsymbol{r}^w\|_2 \leq M \sum_{l=0}^{L} \vartheta_l C_1 N^{5/4} e^{-C_2 N}.$$

We omit the proof due to limited space. The dimension of $\mathcal{S}_{\widehat{\boldsymbol{D}}}$ (the column space of $\widehat{\boldsymbol{D}}$) is J^w, which is smaller than the dimension of $\mathcal{S}_{\boldsymbol{D}}$ by a factor of M. The advantage of this smaller dimension is that the projection operator $\boldsymbol{P}_{\widehat{\boldsymbol{D}}}$ has less effect on the target return $\boldsymbol{r}^t$ than $\boldsymbol{P}_{\boldsymbol{D}}$. We give an example to illustrate this. We simulate one line target of length 0.5m located at $(x, y) = (-0.29\text{m}, 5.38\text{m})$, with complex reflectivity of 5. An $M = 15$-element synthetic linear aperture (located along the x-axis) with interlement spacing of $\frac{4}{M}$m is used. A stepped-frequency signal consisting of $N = 101$ frequencies from 1GHz to 3GHz is utilized to obtain measurements. A front wall is located at $y = 3.13$m, i.e., 3.13m away from the antennas.

We generate target return according to (5.13). On the basis of (5.12), $L = 5$ wall reverberations are generated equally spaced between the wall and 1.5m behind the wall with $\vartheta_0 = 30$ and $\vartheta_l = \frac{1}{1+l}\vartheta_0$ for all $l = 1, \ldots, L$. We have $N(t_L - t_0)\Delta F = 20.2 \approx 20$. Figure 5.1(a-b) respectively show the ability of $\boldsymbol{D}$ (and $\widehat{\boldsymbol{D}}$) to capture the energy in the wall return $\boldsymbol{r}^w$ through the quantification $\text{SNR}_1 = 20\log_{10}(\frac{\|\boldsymbol{r}^w\|_2}{\|\boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^w\|_2})$dB and avoid the target return $\boldsymbol{r}^t$ through the quantification $\text{SNR}_2 = 20\log_{10}(\frac{\|\boldsymbol{r}^t\|_2}{\|\boldsymbol{r}^t - \boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^t\|_2})$ with various J^w near 20. As can be observed, though $\boldsymbol{D}$ and $\widehat{\boldsymbol{D}}$ capture the same energy in the wall return, $\widehat{\boldsymbol{D}}$ captures less energy in the target return.

Also, as anticipated, both $\boldsymbol{D}$ and $\widehat{\boldsymbol{D}}$ may capture non-negligible energy from the target return if we choose J^w too large, whereas choosing J^w too small results in a dictionary that cannot capture the wall return completely. There is a tradeoff between cancelling the wall return and preserving the target return. By changing J^w, we can balance this tradeoff. In general $J^w \approx N(t_L - t_0)\Delta F$ is recommended for most applications.

Wall clutter mitigation

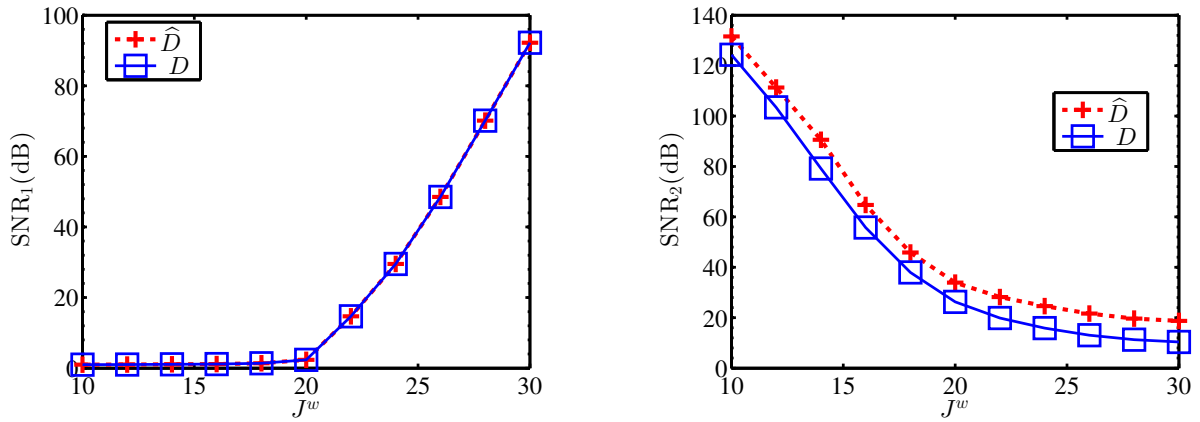


Figure 5.1: (a) SNR_1 captured for wall return as a function of J^w; (b) SNR_2 captured for target return as a function of J^w.

Based on the discussion above, one could mitigate wall clutter antenna-by-antenna by computing [145]

$$\widetilde{\boldsymbol{y}}_m := \boldsymbol{P}_{\boldsymbol{D}_m}\boldsymbol{y}_m = \boldsymbol{P}_{\boldsymbol{D}_m}\boldsymbol{r}_m^w + \boldsymbol{P}_{\boldsymbol{D}_m}\boldsymbol{r}_m^t.$$

Since $\boldsymbol{P}_{\boldsymbol{D}_m}\boldsymbol{r}_m^w \approx 0$, we get $\widetilde{\boldsymbol{y}}_m \approx \boldsymbol{P}_{\boldsymbol{D}_m}\boldsymbol{r}_m^t$. The processed measurements could then be written as

$$\widetilde{\boldsymbol{y}} = \boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^w + \boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^t \approx \boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^t.$$

Alternatively, one could mitigate the wall clutter jointly by computing

$$\widehat{\boldsymbol{y}} = \boldsymbol{P}_{\widehat{\boldsymbol{D}}}\boldsymbol{y} = \boldsymbol{P}_{\widehat{\boldsymbol{D}}}\boldsymbol{r}^w + \boldsymbol{P}_{\widehat{\boldsymbol{D}}}\boldsymbol{r}^t \approx \boldsymbol{P}_{\boldsymbol{D}}\boldsymbol{r}^t.$$

Since $\boldsymbol{P}_{\widehat{\boldsymbol{D}}}$ has less effect on the target return $\boldsymbol{r}^t$ than $\boldsymbol{P}_{\boldsymbol{D}}$, we adopt $\boldsymbol{P}_{\widehat{\boldsymbol{D}}}$ for mitigating the wall return.

5.3.3 Target Return Subspace

The dimensionality of target return subspaces

The target return observed by the m-th antenna corresponding to the k-th target can be rewritten as $\boldsymbol{r}_{k,m}^t[n] = \int_{\tau_{k,m}^{\min}}^{\tau_{k,m}^{\max}} \sigma_k(\tau)e^{-j2\pi f_0\tau}e^{-j2\pi n\tau\Delta F}d\tau$. This indicates that $\boldsymbol{r}_{k,m}^t$ can be viewed

as a linear combination of sampled exponentials $\boldsymbol{e}_f$ with $f \in [-\tau_{k,m}^{\max}\Delta F, -\tau_{k,m}^{\min}\Delta F]$. Thus, from Theorem 5.8, we expect this vector to approximately live within a low-dimensional subspace spanned by certain modulated DPSS vectors. Define

$$\boldsymbol{\Psi}_{k,m} := [\boldsymbol{E}_{-\Delta F(\tau_{k,m}^{\min}+\tau_{k,m}^{\max})/2}\boldsymbol{S}_{N,\Delta F(\tau_{k,m}^{\max}-\tau_{k,m}^{\min})/2}]_{J_{k,m}^t}$$

for some value of $J_{k,m}^t \in \{1, 2, \ldots, N\}$.

Theorem 5.9. *[148] Fix $\epsilon \in (0, 1)$. Choose $J_{k,m}^t = N(\tau_{k,m}^{max} - \tau_{k,m}^{min})\Delta F(1 + \epsilon)$. Then there exist constants C_1, C_2 and an integer N_0 such that for all $N \geq N_0$*

$$\|\boldsymbol{P}_{\boldsymbol{\Psi}_{k,m}}\boldsymbol{r}_{k,m}^t\|_2 \leq \sqrt{\int_{\tau_{k,m}^{\min}}^{\tau_{k,m}^{\max}} \sigma_k^2(\tau)d\tau}C_1N^{5/4}e^{-C_2N}.$$

The above result indicates that $\boldsymbol{r}_{k,m}^t$ is approximately contained within the column space of $\boldsymbol{\Psi}_{k,m}$. Define

$$\boldsymbol{\Psi}_k := \text{diag}(\boldsymbol{\Psi}_{0,k}, \ldots, \boldsymbol{\Psi}_{M-1,k}). \tag{5.15}$$

Let $\overline{\boldsymbol{r}}_k^t := [(\boldsymbol{r}_{k,0}^t)^H \ \cdots \ (\boldsymbol{r}_{k,M-1}^t)^H]^H$ denote the joint target return (across all antennas) corresponding to the k-th target.

Corollary 5.7. *Fix $\epsilon \in (0, 1)$. Choose $J_{k,m}^t = N(\tau_{k,m}^{max} - \tau_{k,m}^{min})\Delta F(1 + \epsilon)$. Then there exist constants C_1, C_2 and an integer N_0 such that for all $N \geq N_0$*

$$\|\boldsymbol{P}_{\boldsymbol{\Psi}_k}\overline{\boldsymbol{r}}_k^t\|_2 \leq \sum_{m=0}^{M-1}\sqrt{\int_{\tau_{k,m}^{\min}}^{\tau_{k,m}^{\max}} \sigma_k^2(\tau)d\tau}C_1N^{5/4}e^{-C_2N}.$$

This result follows directly from Theorem 5.9. In words, the target return $\overline{\boldsymbol{r}}_k^t$ lives approximately within $\mathcal{S}_{\boldsymbol{\Psi}_k}$, the subspace spanned by the columns of $\boldsymbol{\Psi}_k$. The dimension of $\mathcal{S}_{\boldsymbol{\Psi}_k}$ is $J_k^t := \sum_{m=0}^{M-1} J_{k,m}^t$.

Now, similar to the case of the wall return, if we utilize the fact that $\{\boldsymbol{r}_{k,m}^t\}_m$ correspond to the same target, we expect that the joint target return $\overline{\boldsymbol{r}}_k^t$ can be captured using a subspace

with dimension much smaller than $\mathcal{S}_{\boldsymbol{\Psi}_k}$. Divide the k-th target uniformly into P points and construct $\boldsymbol{G}_{k,m} \in \mathbb{C}^{N\times P}$ with entries given by

$$\boldsymbol{G}_{k,m}[n,p] := e^{-j2\pi f_n \tau_{p,m}} \tag{5.16}$$

for $n \in [N]$ and $p \in [P]$. Here $\tau_{p,m}$ denotes the two-way travel time between the p-th point position and the m-th transceiver. One can approximate $\boldsymbol{r}_{k,m}^t$ as a linear combination of the columns of $G_{k,m}$. In fact, an approximate method to generate the modulated DPSS basis $\boldsymbol{\Psi}_{k,m}$ (whose columns are also the eigenvectors of the covariance matrix of a randomly chosen sinusoid in the frequency band of interest, see [32]) is by computing the left singular vectors of $\boldsymbol{G}_{k,m}$.

Choosing P sufficiently large, the matrix $\boldsymbol{G}_{k,m}$ is approximately low rank with effective rank $\approx N(\tau_{k,m}^{\max} - \tau_{k,m}^{\min})\Delta F$. Arrange $\{\boldsymbol{G}_{k,m}\}_{m\in[M]}$ as

$$\boldsymbol{G}_k := [\boldsymbol{G}_{k,0}^H \; \boldsymbol{G}_{k,1}^H \; \cdots \; \boldsymbol{G}_{k,M-1}^H]^H. \tag{5.17}$$

Now, one can approximate $\overline{\boldsymbol{r}}_k^t$ as a linear combination of the columns of $G_{k,m}$. The effective rank of $\boldsymbol{G}_k$ is upper bounded by $\approx \sum_{m=0}^{M-1} N(\tau_{k,m}^{\max} - \tau_{k,m}^{\min})\Delta F$ and lower bounded by $\max_m N(\tau_{k,m}^{\max} - \tau_{k,m}^{\min})\Delta F$. We use an example to illustrate the low rank structure in both $\boldsymbol{G}_{k,m}$ and $\boldsymbol{G}_k$. With the same setup to that in Section 5.3.2, we set $P = 50$. Figure 5.2(a-b) display the singular values of $\boldsymbol{G}_{1,0}$ and $\boldsymbol{G}_1$, respectively. We observe that the effective rank of $\boldsymbol{G}_1$ is only slightly larger than $\boldsymbol{G}_{1,0}$. Here $N(\tau_{1,0}^{\max} - \tau_{1,0}^{\min})\Delta F = 2.04 \approx 2$. Although not shown in the plot, we also note that the effective ranks of $\boldsymbol{G}_{k,m}$ and $\boldsymbol{G}_k$ both scale proportionally with the size of the target.

Let $\boldsymbol{G}_{k,m} = \boldsymbol{U}_{k,m}\boldsymbol{\Sigma}_{k,m}\boldsymbol{V}_{k,m}^H$ be an SVD of $\boldsymbol{G}_{k,m}$, where $\boldsymbol{\Sigma}_{k,m}$ is a diagonal matrix with singular values $\gamma_{k,m}^{(p)}$ (which are arranged in non-increasing order, i.e., $\gamma_{k,m}^{(0)} \geq \gamma_{k,m}^{(1)} \geq \cdots$) along its diagonal. For given $0 < \beta < 1$, define $\overline{J}_{k,m}^t$ as the number of singular values that are greater than $\beta\gamma_{k,m}^{(0)}$, i.e., $\overline{J}_{k,m}^t := \#\{p, \gamma_{k,m}^{(p)} \geq \beta\gamma_{k,m}^{(0)}\}$. Define $\overline{J}_k^t := \sum_{m=0}^{M-1} \overline{J}_{k,m}^t$. Similarly, let $\boldsymbol{G}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^H$ be an SVD of $\boldsymbol{G}_k$, where $\boldsymbol{\Sigma}_k$ is a diagonal matrix with singular values $\gamma_k^{(p)}$

(which are arranged in non-increasing order) along its diagonal. Let $\widehat{J}_k^t$ denote the number of singular values that are greater than $\beta\gamma_k^{(0)}$. Define

$$\overline{\boldsymbol{\Psi}}_k := \mathrm{diag}([\boldsymbol{U}_{k,0}]_{\overline{J}_{k,0}^t} \ \cdots \ [\boldsymbol{U}_{k,M-1}]_{\overline{J}_{k,M-1}^t}), \ \widehat{\boldsymbol{\Psi}}_k := [\boldsymbol{U}_k]_{\widehat{J}_k^t},$$

where $[\boldsymbol{U}_{k,m}]_{\overline{J}_{k,m}^t}$ is obtained by taking the first $\overline{J}_{k,m}^t$ columns of $\boldsymbol{U}_{k,m}$. Similar notation holds for $[\boldsymbol{U}_k]_{\widehat{J}_k^t}$.

We add one more line target of length 0.5m located at $(1.55\text{m}, 6.38\text{m})$, with relative complex reflectivity of 3. Figure 5.2(c) shows J_1^t, $\overline{J}_1^t$ and $\widehat{J}_1^t$ for various β. Here set $J_{k,m}^t = \overline{J}_{k,m}^t$ and thus $J_1^t = \overline{J}_1^t$. We observe that the effective dimensionality of the first target return subspace is much smaller when we consider the antennas jointly. Figure 5.2(d-e) respectively show the ability of $\boldsymbol{\Psi}_1$ (and $\overline{\boldsymbol{\Psi}}_1$, $\widehat{\boldsymbol{\Psi}}_1$) to capture the energy in the first target return $\overline{\boldsymbol{r}}_1^t$ through the quantification $\mathrm{SNR}_1 = 20\log_{10}(\frac{\|\overline{\boldsymbol{r}}_1^t\|_2}{\left\|\boldsymbol{P}_{\boldsymbol{\Psi}_1}\overline{\boldsymbol{r}}_1^t\right\|_2})\text{dB}$ and to avoid the second target return $\overline{\boldsymbol{r}}_2^t$ through the quantification $\mathrm{SNR}_2 = 20\log_{10}(\frac{\|\overline{\boldsymbol{r}}_2^t\|_2}{\left\|\overline{\boldsymbol{r}}_2^t - \boldsymbol{P}_{\boldsymbol{\Psi}_1}\overline{\boldsymbol{r}}_2^t\right\|_2})$ with various β. As can be seen, $\boldsymbol{\Psi}_1, \overline{\boldsymbol{\Psi}}_1$ and $\widehat{\boldsymbol{\Psi}}_1$ have almost the same ability to represent the first target return $\overline{\boldsymbol{r}}_1^t$. However, compared to $\boldsymbol{\Psi}_1$ and $\overline{\boldsymbol{\Psi}}_1$, $\widehat{\boldsymbol{\Psi}}_1$ captures less energy in the second target return $\overline{\boldsymbol{r}}_2^t$. This advantage owes to the fact that $\widehat{\boldsymbol{\Psi}}_1$ has a much smaller number of columns.

Target detection

Following the general approach for radar imaging [2], the target space is divided uniformly into a grid of $L_x \times L_y$ pixels. We arrange the pixels of the image into an $L_xL_y \times 1$ vector $\boldsymbol{\alpha}$. Define $\boldsymbol{\Theta}_m \in \mathbb{C}^{N \times L_xL_y}$ with entries given by $\boldsymbol{\Theta}_m[n,q] := e^{-j2\pi f_n \tau_{q,m}}$ for $n \in [N]$ and $q \in [L_xL_y - 1]$. Here $\tau_{q,m}$ denotes the two-way travel time between the q-th grid and the m-th transceiver. The target return can be written as $\boldsymbol{r}_m^t = \boldsymbol{\Theta}_m\boldsymbol{\alpha}$ if the targets are points and located precisely on the grid. Define $\boldsymbol{\Theta} := [\boldsymbol{\Theta}_0^H \ \cdots \ \boldsymbol{\Theta}_{M-1}^H]^H$.

In order to detect and localize the non-point targets, we [146, 145] modify the iterative, greedy matching pursuit (MP) algorithm [93] so that the energy of exponentials with two-way traveling time close to that of each selected point is cancelled by using a modulated DPSS basis. To account for and cancel off-grid target return, for each $q \in [L_xL_y - 1]$, we generate

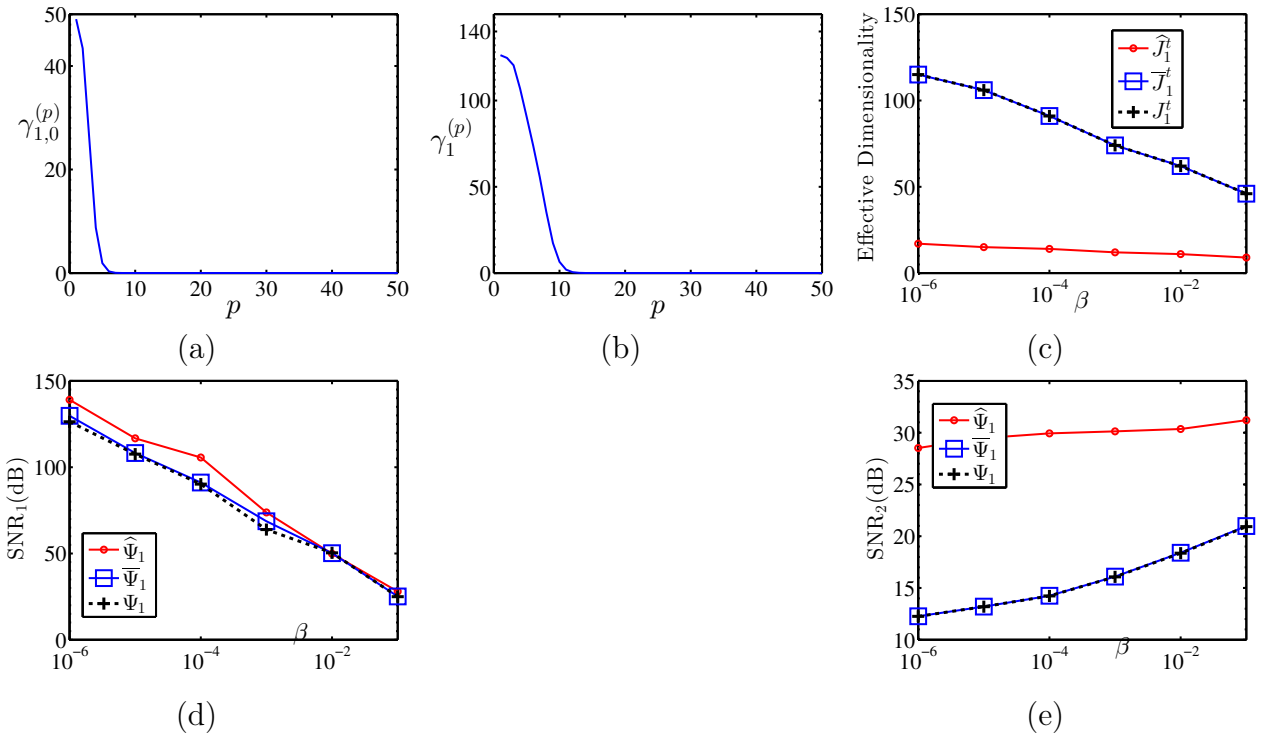


Figure 5.2: (a) The singular values of $\boldsymbol{G}_{1,0}$; (b) the singular values of $\boldsymbol{G}_1$; (c) the effective dimensionality against β; (d) SNR_1 captured for the first target return against β; (e) SNR_2 captured for the second target return against β.

$\overline{\boldsymbol{\Psi}}_q$ and $\widehat{\boldsymbol{\Psi}}_q$ by uniformly dividing a region centered at grid point q with size $R_x\text{m} \times R_y\text{m}$ into $F_x \times F_y$ points, constructing $\boldsymbol{G}_{q,m}$ as defined in (5.16) with $P = F_xF_y$, constructing $\boldsymbol{G}_q$ as defined in (5.17), and finally computing the left singular vectors of $\boldsymbol{G}_{q,m}$ and $\boldsymbol{G}_q$. Throughout the simulations, we choose $R_x = 1$, $R_y = 0.3$, $F_x = 12$ and $F_y = 6$. $\boldsymbol{\Psi}_q$ is also generated for this region according (5.15). The full subspace-based MP algorithm for target detection is shown in Algorithm 1. As shown in the merge step, in each iteration when we pick one pixel in the grid, we also choose its neighbors (from two pixels to the left to two pixels to the right). We note that this step (adding its neighbors) is used only to improve the imaging result, but has no effect in detection. The size of the neighbors can be adapted to the particular application.

Algorithm 1 Subspace-based Matching Pursuit.

Require: $\boldsymbol{\Theta}$ with columns $\boldsymbol{\theta}_j$, $\widehat{\boldsymbol{y}}$, number of iterations I
Ensure: $\boldsymbol{r}^0 = \widehat{\boldsymbol{y}}, \widehat{\boldsymbol{\alpha}} = 0, i = 0, \Lambda^0 = \varnothing$
1: **while** $i < I$ **do**
2: identify: $j_0 = \arg\max_j |\boldsymbol{\theta}_j^H \boldsymbol{r}^i| / ||\boldsymbol{\theta}_j||_2$
3: merge: $\Lambda^{i+1} = \Lambda^i \cup \{j_0 - 2, j_0 - 1, j_0, j_0 + 1, j_0 + 2\}$
4: update: $\boldsymbol{r}^{i+1} = \boldsymbol{P}_{\boldsymbol{\Psi}_{j_0}} \boldsymbol{r}^i$ (or $\boldsymbol{P}_{\overline{\boldsymbol{\Psi}}_{j_0}} \boldsymbol{r}^i$, $\boldsymbol{P}_{\widehat{\boldsymbol{\Psi}}_{j_0}} \boldsymbol{r}^i$)
$i = i + 1$
5: **end while**
6: **return** $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\Theta}_\Lambda^\dagger \widehat{\boldsymbol{y}}$

5.3.4 Simulations

With the same setup to that in Section 5.3.2, we simulate eight line targets of length 0.5m as listed in Table 5.1. The 4m $\times$ 5.5m region centered at $(0\text{m}, 4.75\text{m})$ is chosen to be imaged, and it is divided into a grid of 33×77 pixels. The number J^w for the bandpass modulated DPSS dictionary $\boldsymbol{D}$ is chosen to be 30. We choose $\beta = 10\%$.

Table 5.1: The location and reflectivity of the targets

k	1	2	3	4	5	6	7	8
x(m)	-0.29	1.55	-1.69	-1.78	1.67	0.29	1	-1.3
y(m)	5.38	6.38	6.58	4.93	5.03	6.6	4.97	3.63
σ	5	3	2	1	1	1	1	1

Figure 5.3(b-d) respectively display the target reconstruction result with 8 iterations of the subspace-based MP algorithm involving $\boldsymbol{\Psi}$, $\overline{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Psi}}$ (which means we use $\boldsymbol{\Psi}_{j_0}$, $\overline{\boldsymbol{\Psi}}_{j_0}$ and $\widehat{\boldsymbol{\Psi}}_{j_0}$ respectively in Algorithm 1). We note that the wall clutter can be captured well by $\widehat{\boldsymbol{D}}$ and due to the limited space, we only show the region containing the targets in Figure 5.3. Clearly, we observe that the algorithm with $\widehat{\boldsymbol{\Psi}}$ can find the second and the fifth targets which are very close to each other, while the algorithms with $\boldsymbol{\Psi}$ and $\overline{\boldsymbol{\Psi}}$ miss the fifth target.

As we have explained, the wall and target return subspaces have much smaller dimensionality when we consider them jointly across all antenna elements than separately for each antenna. This experiment demonstrates the advantage of using the joint target subspace in target detection.

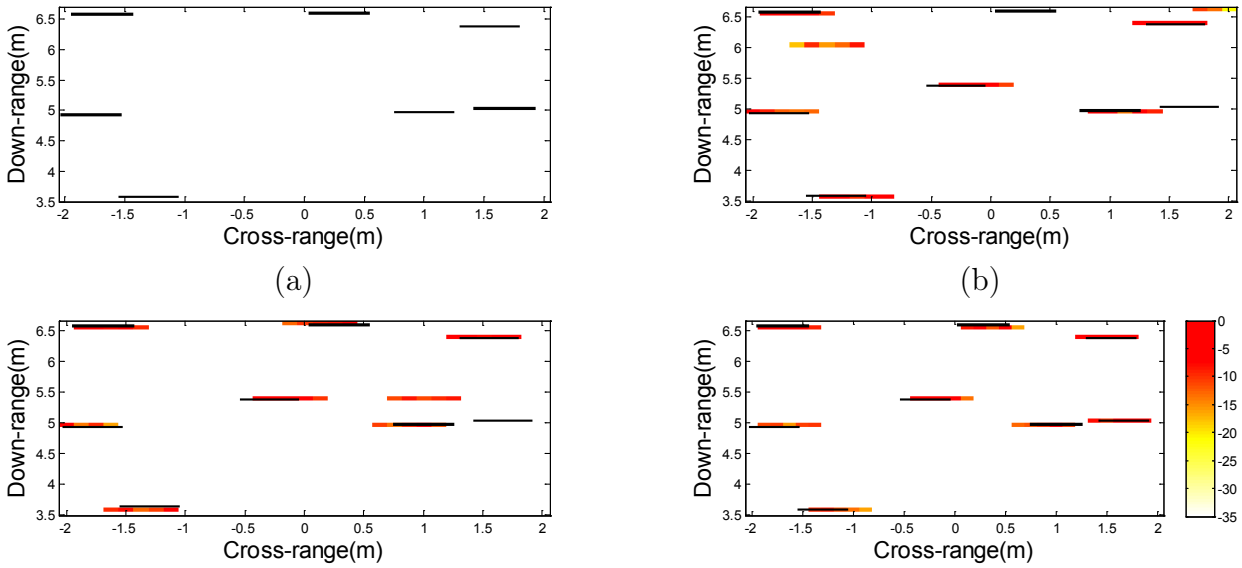


Figure 5.3: Illustration of (a) the true target locations; and wall mitigation with $\widehat{\boldsymbol{D}}$ and target detection with subspace-based MP algorithm involving (b) $\boldsymbol{\Psi}$; (c) $\overline{\boldsymbol{\Psi}}$; (d) $\widehat{\boldsymbol{\Psi}}$.

CHAPTER 6

ON THE ASYMPTOTIC EQUIVALENCE OF CIRCULANT AND TOEPLITZ MATRICES

Motivated by the facts that Toeplitz matrices appear naturally for parameterized subspace models (for example, the prolate matrix defined in (2.11) is Toeplitz) and that the number of large eigenvalues of Toeplitz matrices reveal the required dimensionality of a potential subspace for representing signals obeying certain parameterized subspace model (see Section 2.7 and Section 7.1), in this chapter[12], we study a fast way to approximately compute the spectrum of Toeplitz matrices $\boldsymbol{H}_N$ (see (2.22) for the formal definition of Toepltiz matrices). Rather than invoking the Szegő's theorem which requires the information of the generating function $\widetilde{h}$ (see Section 2.8), we construct circulant matrices that are asymptotically equivalent to the Toepltiz matrices. We then provide conditions under which the asymptotic equivalence of the matrices implies the individual asymptotic convergence of the eigenvalues. Our results suggest that instead of directly computing the eigenvalues of a Toeplitz matrix, one can compute a fast spectrum approximation using the FFT.

6.1 Motivation

Despite the power of Szegő's theorem (see Section 2.8), in many scenarios (such as certain coding and filtering applications [54, 104]), one may only have access to $\boldsymbol{H}_N$ and not its generating function $\widetilde{h}$. In such cases, it is still desirable to have practical and efficiently computable estimates of individual eigenvalues of $\boldsymbol{H}_N$. We elaborate on two example applications below.

i. **Estimating the condition number of a positive-definite Toeplitz matrix.** The linear system $\boldsymbol{H}_N\boldsymbol{y} = \boldsymbol{b}$ arises naturally in many signal processing and estimation problems

[12]This work is in collaboration with Michael B. Wakin [149].

such as linear prediction [91, 73]. The condition number $\kappa(\boldsymbol{H}_N)$ of the Toeplitz matrix $\boldsymbol{H}_N$ is important when solving such systems. For example, the speed of solving such linear systems via the widely used conjugate gradient method is determined by the condition number: the larger $\kappa(\boldsymbol{H}_N)$, the slower convergence of the algorithm. In case of large $\kappa(\boldsymbol{H}_N)$, preconditioning can be applied to ensure fast convergence. Thus estimating the smallest and largest eigenvalues of a symmetric positive-definite Toeplitz matrix (such as the covariance matrix of a stationary random process) is of considerable interest [36, 84].

ii. **Spectrum sensing algorithm for cognitive radio.** Spectrum sensing is a fundamental task in cognitive ratio, which aims to best utilize the available spectrum by identifying unoccupied bands [97, 61, 142]. Zeng and Ling [142] have proposed spectrum sensing methods for cognitive radio based on the eigenvalues of a Toeplitz covariance matrix. These eigenvalue-based algorithms overcome the noise uncertainty problem which exists in alternative methods based on energy detection.

Aside from the above applications, approximate and efficiently computable eigenvalue estimates can also be used as the starting point for numerical algorithms that iteratively compute eigenvalues with high precision.

6.2 Circulant Approximations

In this section, we consider estimates for the eigenvalues of a Toeplitz matrix that are obtained through a two-step process:

1. Transform the Toeplitz matrix into a circulant matrix using a certain procedure described below.

2. Compute the eigenvalues of the circulant matrix.

Both of these steps can be performed efficiently; in particular, the eigenvalues of an $N \times N$ circulant matrix can be computed in $O(N \log N)$ time using the fast Fourier transform (FFT). The individual eigenvalues of the circulant matrix approximate those of the Toeplitz matrix.

We study the quality of this approximation.

An $N \times N$ circulant matrix $\boldsymbol{C}_N$ is a special Toeplitz matrix of the form

$$\boldsymbol{C}_N = \begin{bmatrix} c[0] & c[1] & c[2] & \dots & c[N-1] \\ c[N-1] & c[0] & c[1] & & \\ c[N-2] & c[N-1] & c[0] & & \vdots \\ \vdots & & & \ddots & \\ c[1] & & \cdots & & c[0] \end{bmatrix}.$$

Circulant matrices arise naturally in applications involving the discrete Fourier transform (DFT) [104]; in particular, any circulant matrix can be unitarily diagonalized using the DFT matrix. Circulant matrices offer a nontrivial but simple set of objects that can be used for problems involving Toeplitz matrices. For example, the product $\boldsymbol{H}_N\boldsymbol{x}$ can be computed in $O(N \log N)$ time by embedding $\boldsymbol{H}_N$ into a $(2N-1)\times(2N-1)$ circulant matrix and using the FFT to perform matrix-vector multiplication. Also Gray [54, 55] showed that Toeplitz and circulant matrices are asymptotically equivalent in a certain sense; this implies that their eigenvalues have similar *collective behavior*. See Section 6.3.1 for formal definitions. Finally, we note that circulant matrices have been used as preconditioners [122, 22] of Toeplitz matrices in iterative methods for solving linear systems of the form $\boldsymbol{H}_N\boldsymbol{y} = \boldsymbol{b}$.

We consider estimates for the eigenvalues of a Toeplitz matrix obtained from a well-constructed circulant matrix. The eigenvalues of the circulant matrix can be computed efficiently without constructing the whole matrix; one merely applies the FFT to the first row of the matrix. We do *not* provide new circulant approximations to Toeplitz matrices in this paper; rather we sharpen the analysis on the asymptotic equivalence of Toeplitz and certain circulant matrices [54, 104, 55] by establishing results in terms of *individual eigenvalues* rather than collective behavior. To the best of our knowledge, this is the first work that provides guarantees for asymptotic equivalence in terms of individual eigenvalues.

We consider the following circulant approximations that have been widely used in information theory and applied mathematics.

Type I: $\widetilde{\boldsymbol{C}}_N$

Bogoya et al. [12] proved that the samples of the symbol $\widetilde{h}$ are the main asymptotic terms of the eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$. Given only $\boldsymbol{H}_N$, one practical strategy for estimating the eigenvalues is to first approximate $\widetilde{h}$ by the $(N-1)^{\text{th}}$ partial Fourier sum $S_{N-1}(f) = \sum_{k=-(N-1)}^{N-1} h[k]e^{j2\pi fk}$. Then construct a circulant matrix whose eigenvalues are samples of $S_{N-1}(f)$, i.e., $S_{N-1}(\frac{l}{N})$. We let $\widetilde{\boldsymbol{C}}_N$ denote the corresponding circulant matrix, whose top row $(\widetilde{c}[0], \widetilde{c}[1], \ldots, \widetilde{c}[N-1])$ can be obtained as

$$\begin{aligned}\widetilde{c}[k] &= \frac{1}{N}\sum_{n=0}^{N-1} S_{N-1}(\frac{2\pi n}{N})e^{j2\pi kn/N} = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{k'=-(N-1)}^{N-1} h[k']e^{j2\pi(k+k')n/N} \\ &= \sum_{k'=-(N-1)}^{N-1} h[k']\left(\sum_{n=0}^{N-1}\frac{1}{N}e^{j2\pi(k+k')n/N}\right) = \begin{cases} h[0], & k=0, \\ h[-k]+h[N-k], & k=1,2,\ldots,N-1, \end{cases}\end{aligned}$$

where the last line utilizes the fact

$$\sum_{n=0}^{N-1}\frac{1}{N}e^{j2\pi(k+k')n/N} = \begin{cases} 1, & \text{mod}(k+k', N) = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Type II: $\widehat{\boldsymbol{C}}_N$

Following the same strategy, we first compute the $\left(\lfloor\frac{N-1}{2}\rfloor\right)^{\text{th}}$ partial Fourier sum

$$S_{\lfloor\frac{N-1}{2}\rfloor}(f) = \sum_{k=-\lfloor\frac{N-1}{2}\rfloor}^{\lfloor\frac{N-1}{2}\rfloor} h[k]e^{j2\pi fk}.$$

Let $\widehat{\boldsymbol{C}}_N$ denote the $N\times N$ circulant matrix whose eigenvalues are samples of $S_{\lfloor\frac{N-1}{2}\rfloor}(f)$, i.e, $S_{\lfloor\frac{N-1}{2}\rfloor}(\frac{l}{N})$. With simple manipulations, the top row $(\widehat{c}[0], \widehat{c}[1], \ldots, \widehat{c}[N-1])$ of $\widehat{\boldsymbol{C}}_N$ is given by

$$\widehat{c}[k] = \begin{cases} h[-k], & 0 \le k \le \lfloor \frac{N-1}{2} \rfloor, \\ h[N-k], & \lceil \frac{N+1}{2} \rceil \le k < N, \\ 0, & k = N/2, \end{cases}$$

when N is even, and

$$\widehat{c}[k] = \begin{cases} h[-k], & 0 \le k \le \lfloor \frac{N-1}{2} \rfloor, \\ h[N-k], & \lceil \frac{N+1}{2} \rceil \le k < N, \end{cases}$$

when N is odd.

Strang [122] first employed such circulant matrices as preconditioners to speed up the convergence of iterative methods for solving Toeplitz linear systems. This approach is quite simple. The underlying idea is that the sequence $h[k]$ usually decays quickly as k grows large, and thus we keep the largest part of the Toeplitz matrix and fill in the remaining part to form a circulant approximation.

Type III: $\overline{\boldsymbol{C}}_N$

In the Fourier analysis literature, it is known that Cesàro sum has rather better convergence than the partial Fourier sum [77]. The N^{th} Cesàro sum is defined as

$$\sigma_N(f) = \frac{\sum_{n=0}^{N-1} S_n(f)}{N}.$$

We use $\overline{\boldsymbol{C}}_N$ to denote the $N \times N$ circulant matrix whose eigenvalues are samples of $\sigma_N(f)$, i.e., $\sigma_N(\frac{l}{N})$. The top row $(\overline{c}[0], \overline{c}[1], \ldots, \overline{c}[N-1])$ of $\overline{\boldsymbol{C}}_N$ can be obtained as follows

$$\begin{aligned} \overline{c}[k] &= \frac{1}{N} \sum_{l=0}^{N-1} \sigma_N(\frac{l}{N}) e^{j2\pi kl/N} = \frac{1}{N} \sum_{l=0}^{N-1} \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k'=-n}^{n} h[k'] e^{j2\pi l(k+k')/N} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k'=-n}^{n} \left(h[k'] \sum_{l=0}^{N-1} \frac{1}{N} e^{j2\pi l(k+k')/N} \right) = \frac{1}{N} \left((N-k)h[-k] + kh[N-k] \right). \end{aligned}$$

Pearl [104] first analyzed such a circulant approximation and its applications in coding and filtering. The same circulant approximation (referred to as an optimal preconditioner)

was also proposed by Chan [22]. The optimal preconditioner is the solution to the following optimization problem

$$\text{minimize } \|\boldsymbol{C}_N - \boldsymbol{H}_N\|_F$$

over all $N \times N$ circulant matrices. One can verify that $\overline{\boldsymbol{C}}_N$ is the solution to the above problem.

6.3 Asymptotic Equivalence of Circulant and Toeplitz Matrices

We first give out the notion of asymptotic equivalence of two set of matrices.

6.3.1 Asymptotically Equivalent Matrices

We begin with the notion of equal distribution of two real sequences, using a definition attributed to Weyl [56].

Definition 6.1. *[56] Assume that the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are absolutely bounded, i.e., there exist a, b such that $a \le u_{N,l} \le b$ and $a \le v_{N,l} \le b$ for all $l \in [N]$ and $N \in \mathbb{N}$. Then $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are equally distributed if*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right) = 0.$$

for every continuous function ϑ on $[a, b]$.

The asymptotic equivalence of two sequences of matrices is defined as follows.

Definition 6.2. *[54, 55] Two sequences of $N \times N$ matrices $\{\boldsymbol{A}_N\}$ and $\{\boldsymbol{B}_N\}$ (where $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ denote $N \times N$ matrices) are said to be asymptotically equivalent if*

$$\lim_{N\to\infty} \frac{\|\boldsymbol{A}_N - \boldsymbol{B}_N\|_F}{\sqrt{N}} = 0$$

and there exists a constant $M < \infty$ such that

$$\|\boldsymbol{A}_N\|_2, \|\boldsymbol{B}_N\|_2 \le M, \quad \forall N \in \mathbb{N}.$$

Following the convention in Gray's monograph [55], we write $\boldsymbol{A}_N \sim \boldsymbol{B}_N$ if $\{\boldsymbol{A}_N\}$ and $\{\boldsymbol{B}_N\}$ are asymptotically equivalent. This kind of asymptotic equivalence is transitive, i.e., if $\boldsymbol{A}_N \sim \boldsymbol{B}_N$ and $\boldsymbol{B}_N \sim \boldsymbol{C}_N$, then $\boldsymbol{A}_N \sim \boldsymbol{C}_N$. Additional properties of $\sim$ can be found in [55]. The following result concerns the asymptotic eigenvalue behavior of asymptotically equivalent Hermitian matrices.

Theorem 6.1. *[55, Theorem 2.4] Let $\{\boldsymbol{A}_N\}$ and $\{\boldsymbol{B}_N\}$ be asymptotically equivalent sequences of Hermitian matrices with eigenvalues $\{\{\lambda_l(\boldsymbol{A}_N)\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{\lambda_l(\boldsymbol{B}_N)\}_{l\in[N]}\}_{N=1}^{\infty}$. Then there exist constants a and b such that*

$$a \leq \lambda_l(\boldsymbol{A}_N), \lambda_l(\boldsymbol{B}_N) \leq b, \ \forall \ l \in [N], N \in \mathbb{N}.$$

Let ϑ be any function continuous on $[a, b]$. We have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(\lambda_l(\boldsymbol{A}_N)) - \vartheta(\lambda_l(\boldsymbol{B}_N))\right) = 0.$$

In light of this theorem, Definition 6.2 can be viewed as the matrix equivalent of Definition 6.1.

6.3.2 Asymptotic Equivalence of Circulant and Toeplitz Matrices

Any circulant matrix $\boldsymbol{C}_N$ is characterized by its top row. Note that

$$\left(\boldsymbol{C}_N \boldsymbol{e}_{(N-l)/N}\right)[k] = \sum_{n=0}^{N-1} c[n] e^{j2\pi(N-l)(k+n)/N} = e^{j2\pi(N-l)k/N} \left(\sum_{n=0}^{N-1} c[n] e^{-j2\pi ln/N}\right),$$

which implies that

$$\boldsymbol{C}_N \boldsymbol{e}_{(N-l)/N} = \left(\sum_{n=0}^{N-1} c[n] e^{-j2\pi ln/N}\right) \boldsymbol{e}_{(N-l)/N}.$$

Thus the normalized DFT basis vectors $\left\{\frac{1}{\sqrt{N}}\boldsymbol{e}_{l/N}\right\}_{l\in[N]}$ are the eigenvectors of any circulant matrix $\boldsymbol{C}_N$, and the corresponding eigenvalues are obtained by taking the DFT of the first

row of $\boldsymbol{C}_N$. Specifically,

$$\lambda_l\left(\boldsymbol{C}_N\right) = \sum_{n=0}^{N-1} c[n] e^{-j2\pi ln/N},$$

which can be computed efficiently via the FFT. We note that $\{\lambda_l\left(\boldsymbol{C}_N\right)\}_{l\in[N]}$ are not necessarily arranged in any particular order; namely, they do not necessarily decrease with l.

For a sequence of Toeplitz matrices $\{\boldsymbol{H}_N\}$ and their respective circulant approximations discussed before, the following result establishes asymptotic equivalence in terms of the collective behaviors of the eigenvalues. As a reminder, we assume throughout this paper that each $\boldsymbol{H}_N$ is Hermitian; this ensures that all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$ are Hermitian as well.

Lemma 6.1. *Suppose that the sequence $h[k]$ is square summable and $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N$ are absolutely bounded*[13] *for all $N \in \mathbb{N}$. Then*

$$\boldsymbol{H}_N \sim \widehat{\boldsymbol{C}}_N \sim \widetilde{\boldsymbol{C}}_N \sim \overline{\boldsymbol{C}}_N,$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(\lambda_l(\boldsymbol{H}_N)) - \vartheta(\lambda_l(\boldsymbol{C}_N))\right) = 0,$$

where ϑ is any continuous function on $[a, b]$ and $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. Here $[a, b]$ is the smallest interval that covers all the eigenvalues of $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$.

The proof is given in Appendix D.5. A stronger result follows simply from the elementary view of Weyl's theory of equal distribution [126], which is presented in Lemma D.6. As a reminder, we do assume that the eigenvalues of each Toeplitz matrix are ordered such that $\lambda_0(\boldsymbol{H}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{H}_N)$.

Lemma 6.2. *Suppose that the sequence $h[k]$ is square summable and $\boldsymbol{H}_N, \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N$ are absolutely bounded. Let $\lambda_l(\boldsymbol{C}_N)$ be permuted that such that $\lambda_{\rho(0)}(\boldsymbol{C}_N) \geq \lambda_{\rho(1)}(\boldsymbol{C}_N) \geq \cdots \geq$*

[13] We say a matrix $\boldsymbol{A}$ is absolutely bounded if its spectral norm (or largest singular value) $\|\boldsymbol{A}\|_2$ is bounded.

$\lambda_{\rho(N-1)}(\boldsymbol{C}_N)$. *Then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\left|\vartheta(\lambda_l(\boldsymbol{H}_N))-\vartheta(\lambda_{\rho(l)}(\boldsymbol{C}_N))\right|=0$$

for every function ϑ *that is continuous on* $[a,b]$ *and* $\boldsymbol{C}_N\in\left\{\widetilde{\boldsymbol{C}}_N,\widehat{\boldsymbol{C}}_N,\overline{\boldsymbol{C}}_N\right\}$. *Here* $[a,b]$ *is the smallest interval that covers all the eigenvalues of* $\boldsymbol{H}_N,\widetilde{\boldsymbol{C}}_N,\widehat{\boldsymbol{C}}_N$, *and* $\overline{\boldsymbol{C}}_N$.

This result follows simply from Lemmas 6.1 and D.6.

6.4 Individual Eigenvalue Estimates

Let $\{\lambda_l(\boldsymbol{C}_N)\}_{l\in[N]}$ denote the eigenvalues of the circulant matrix $\boldsymbol{C}_N$ for all $\boldsymbol{C}_N\in\left\{\widetilde{\boldsymbol{C}}_N,\widehat{\boldsymbol{C}}_N,\overline{\boldsymbol{C}}_N\right\}$. Let $\lambda_l(\boldsymbol{C}_N)$ be permuted that such that $\lambda_{\rho(0)}(\boldsymbol{C}_N)\geq\lambda_{\rho(1)}(\boldsymbol{C}_N)\geq\cdots\geq\lambda_{\rho(N-1)}(\boldsymbol{C}_N)$. As the main contributions of this chapter, in this section, we establish the following results, whose proof is given in Appendix D.

Theorem 6.2. *Suppose that the sequence* $h[k]$ *is absolutely summable. Then*

$$\lim_{N\to\infty}\max_{l\in[N]}\left|\lambda_l(\boldsymbol{H}_N)-\lambda_{\rho(l)}(\boldsymbol{C}_N)\right|=0, \tag{6.1}$$

for all $\boldsymbol{C}_N\in\left\{\widetilde{\boldsymbol{C}}_N,\widehat{\boldsymbol{C}}_N,\overline{\boldsymbol{C}}_N\right\}$.

Theorem 6.2 states that the individual asymptotic convergence of the eigenvalues between the Toeplitz matrices $\boldsymbol{H}_N$ and circulant matrices $\boldsymbol{C}_N\in\left\{\widetilde{\boldsymbol{C}}_N,\widehat{\boldsymbol{C}}_N,\overline{\boldsymbol{C}}_N\right\}$ holds as long as $h[k]$ is absolutely summable. Its proof involves the uniform convergence of a Fourier series and the fact that the equal distribution of two sequences implies individual asymptotic equivalence of two sequences in a certain sense. By utilizing the Sturmian separation theorem [67], we also provide the convergence rate for band Toeplitz matrices as follows.

Theorem 6.3. *Suppose that* $h[k]=0$ *for all* $k>r$, *i.e.,* $\boldsymbol{H}_N$ *is a band Toeplitz matrix when* $N>r$. *Then*

$$\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right| = O(\frac{1}{N}) \tag{6.2}$$

as $N \to \infty$ *for all* $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$.

Utilizing the fact that the Cesàro sum has rather better convergence than the partial Fourier sum, the following result establishes a weaker condition on $h[k]$ for the individual asymptotic convergence of the eigenvalues between $\boldsymbol{H}_N$ and $\overline{\boldsymbol{C}}_N$.

Theorem 6.4. *Suppose that* $h[k]$ *is square summable and* $\widetilde{h} \in L^\infty([0,1])$ *is Riemann integrable and the essential range of* $\widetilde{h}$ *is* $\left[\operatorname{ess\,inf} \widetilde{h}, \operatorname{ess\,sup} \widetilde{h}\right]$, *i.e., the essential range of* $\widetilde{h}$ *is connected. Then*

$$\lim_{N\to\infty} \max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)\right| = 0. \tag{6.3}$$

Note that the sequence $h[k]$ being absolutely summable implies that $h[k]$ is square summable, that $\widetilde{h} \in L^\infty([0,1])$ is Riemann integrable, and that its range is connected. However, the converse of this statement does not hold. We provide an example in the simulation part.

Finally, the following result concerns the convergence of the largest and smallest eigenvalues for more general classes of Toeplitz matrices.

Theorem 6.5. *Suppose that* $\widetilde{h} \in L^\infty([0,1])$ *is Riemann integrable. Then*

$$\begin{aligned}
\lim_{N\to\infty} \lambda_0\left(\boldsymbol{H}_N\right) &= \lim_{N\to\infty} \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,sup} \widetilde{h}, \\
\lim_{N\to\infty} \lambda_{N-1}\left(\boldsymbol{H}_N\right) &= \lim_{N\to\infty} \lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,inf} \widetilde{h}.
\end{aligned}$$

Laudadio et al. [84] summarized several algorithms to estimate the smallest eigenvalue of a symmetric positive-definite Toeplitz matrix. These algorithms need $O(N^2)$ flops. Com-

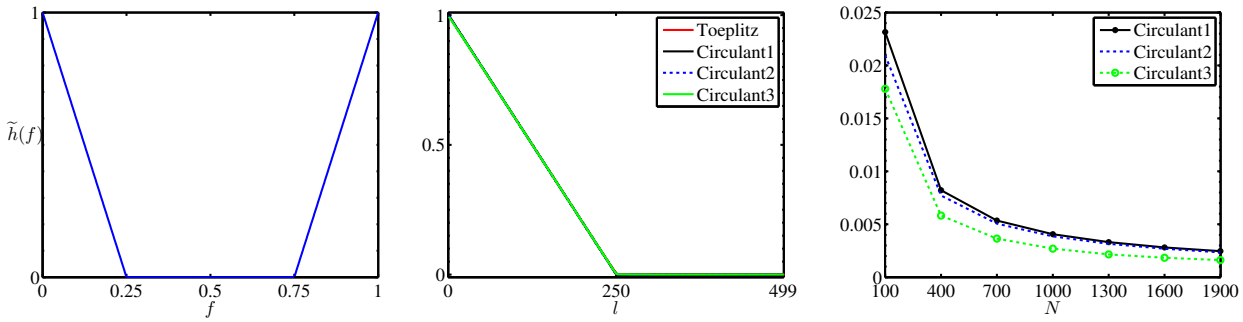
puting $\lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right)$ via the FFT requires $O(N \log N)$ flops, and at the same time, we are guaranteed that $\lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$ is asymptotically equivalent to $\lambda_{N-1}\left(\boldsymbol{H}_N\right)$ by Theorem 6.5.

The above results—characterizing the individual asymptotic convergence of the eigenvalues between Toeplitz and circulant matrices—serve as complements to the literature on asymptotic equivalence that has focused on the collective behavior of the eigenvalues. Before moving on, we briefly review said literature. In [54, 55], Gray showed the asymptotic equivalence of $\{\boldsymbol{H}_N\}$ and $\{\widetilde{\boldsymbol{C}}_N\}$ when the sequence $h[k]$ is absolutely summable. Pearl showed the asymptotic equivalence of $\{\boldsymbol{H}_N\}$ and $\{\overline{\boldsymbol{C}}_N\}$ when the sequence $h[k]$ is square summable and $\boldsymbol{H}_N$ and $\overline{\boldsymbol{C}}_N$ have bounded eigenvalues for all $N \in \mathbb{N}$. The spectrum of the preconditioned matrix $\boldsymbol{C}_N^{-1}\boldsymbol{H}_N$ asymptotically clustering around one was investigated in [19, 21, 20, 130]. Finally, as noted previously, Bogoya et al. [12] studied the *individual* asymptotic behavior of the eigenvalues of Toeplitz matrices by interpreting Szegő's theorem in probabilistic language. Our estimates for the eigenvalues of a Toeplitz matrix differ from [12] in that they are only dependent on the entries of $\boldsymbol{H}_N$ (instead of the symbol $\widetilde{h}(f)$). For our proof, we utilize the same approach of interpreting Szegő's theorem in probabilistic language. However, [12] requires the sequences of the eigenvalues to be strictly inside the range of $\widetilde{h}$, while our work covers more general cases where the sequences of the eigenvalues can be outside of the range of $\widetilde{h}$ as illustrated in Theorem D.1. See also our remark at the end of Section D.3.

6.5 Simulations

In this section, we provide several examples to illustrate our theory. In the legends of Figure 6.1–Figure 6.3, we refer to the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$ as Circulant1, Circulant2, and Circulant3, respectively.

Case I: $h[k] = W\left(\frac{\sin(\pi W k)}{\pi k}\right)^2, W = \frac{1}{4}$



(a) (b) (c)

Figure 6.1: (a) Illustration of a continuous symbol $\widetilde{h}(f)$. (b) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ and the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. Here $N = 500$. (c) A plot of $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ versus the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$.

In our first example, the sequence $h[k]$ is absolutely summable and the corresponding symbol

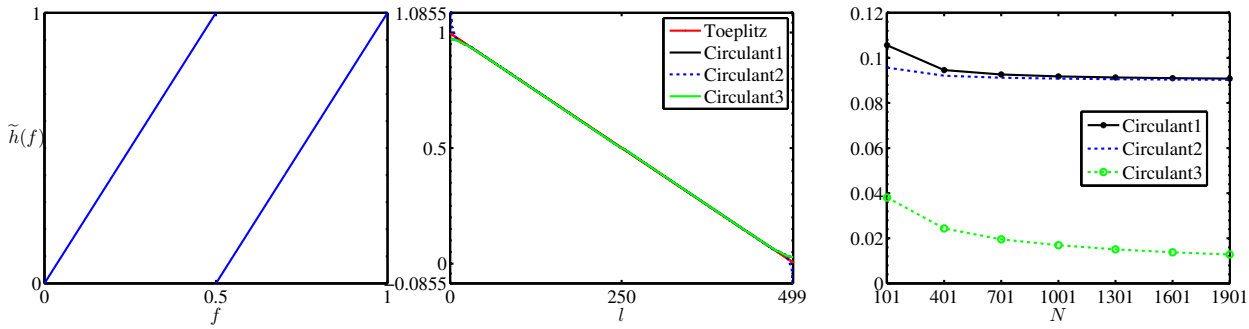
$$\widetilde{h}(f) = \mathrm{tri}(\frac{f}{W}) = \begin{cases} 1 - \frac{f}{W}, & 0 \le f \le W \\ 1 - \frac{1-f}{W}, & 1 - W \le f \le 1 \\ 0, & \text{otherwise} \end{cases}$$

is a triangular signal, which is continuous on $[0, 1]$. Figure 6.1(a) shows $\widetilde{h}$, Figure 6.1(b) shows $\lambda_l(\boldsymbol{H}_N)$, $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$, $\lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 500$, and Figure 6.1(c) shows $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ against the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. As guaranteed by Theorem 6.2, it can be observed in Figure 6.1(c) that the individual asymptotic convergence of eigenvalues holds for all $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$.

Case II: $h[k] = \frac{1+(-1)^k}{j2\pi k}$

In this case, the sequence $h[k]$ is not absolutely summable and the symbol

$$\widetilde{h}(f) = \begin{cases} 2f, & 0 < f \le \frac{1}{2} \\ 2f - 1, & \frac{1}{2} < f \le 1 \end{cases}$$



(a) (b) (c)

Figure 6.2: (a) Illustration of a discontinuous symbol $\widetilde{h}(f)$. (b) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ and the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. Here $N = 500$. (c) A plot of $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ versus the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$.

is not continuous, but its range is connected. Figure 6.2(a) shows $\widetilde{h}$, Figure 6.2(b) shows $\lambda_l(\boldsymbol{H}_N)$, $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$, $\lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 500$, and Figure 6.2(c) shows

$$\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$$

against the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. It is observed from Figure 6.2(c) that the individual asymptotic convergence of the eigenvalues holds for $\overline{\boldsymbol{C}}_N$—as guaranteed by Theorem 6.4—but not for $\widetilde{\boldsymbol{C}}_N$ and $\widehat{\boldsymbol{C}}_N$. Figure 6.2(c) also shows that the errors $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)\right|$ and $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)\right|$ converge to the size of the Gibbs jump (≈ 0.089).

Case III: $h[k] = \frac{\sin(2\pi W k)}{\pi k}, W = \frac{1}{4}$

In this example, the sequence $h[k]$ is not absolutely summable and the symbol

$$\widetilde{h}(f) = \begin{cases} 0, & W < f \leq 1 - W, \\ 1, & \text{otherwise}, \end{cases}$$

is a rectangular window function, which is not continuous and whose range is not connected. Figure 6.3(a) shows $\widetilde{h}$, Figure 6.3(b) shows $\lambda_l(\boldsymbol{H}_N)$, $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$, $\lambda_{\rho(l)}(\widehat{\boldsymbol{C}}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 2048$, and Figure 6.3(c) shows $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ against the dimension

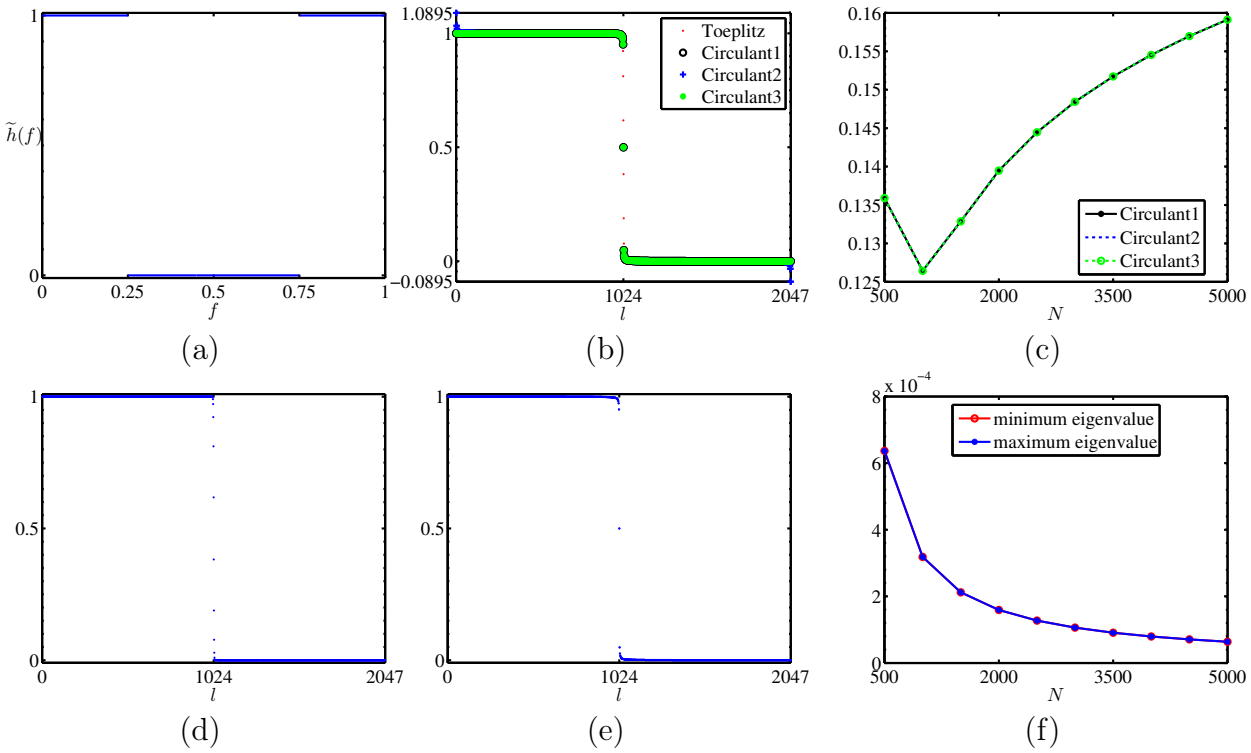


Figure 6.3: (a) Illustration of a discontinuous symbol $\widetilde{h}(f)$ whose range is not connected. (b) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ and the circulant approximations $\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. Here $N = 2048$. (c) A plot of $\max_{l\in[N]} \left|\lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\boldsymbol{C}_N)\right|$ versus the dimension N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. (d) The eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$. (e) The eigenvalues of the circulant matrix $\overline{\boldsymbol{C}}_N$, arranged in decreasing order. (f) A plot of $\left|\lambda_0(\boldsymbol{H}_N) - \lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N)\right|$ and $\left|\lambda_{N-1}(\boldsymbol{H}_N) - \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)\right|$ versus the dimension N.

N for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. Figure 6.3(c) illustrates that the individual asymptotic convergence of eigenvalues does not hold for the circulant matrices $\widetilde{\boldsymbol{C}}_N$, $\widehat{\boldsymbol{C}}_N$, and $\overline{\boldsymbol{C}}_N$. Indeed, the sequence $h[k]$ does not meet the assumptions in either Theorem 6.2 or Theorem 6.4.

Due to the gap (between 0 to 1) in the range of the window function $\widetilde{h}$, the eigenvalues of $\boldsymbol{H}_N$ and $\overline{\boldsymbol{C}}_N$ have different behavior in the transition region. To better illustrate this, Figure 6.3(d) and Figure 6.3(e), respectively, show $\lambda_l(\boldsymbol{H}_N)$ and $\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$ for $N = 2048$. We see that the eigenvalues of the Toeplitz matrix $\boldsymbol{H}_N$ cover the range $[0, 1]$ somewhat uniformly, while the eigenvalues of the $\overline{\boldsymbol{C}}_N$ tend to cluster around 0, 1/2, and 1 (there are none near 1/4 or 3/4). The following result formally explains the transition behavior of the eigenvalues

of $\boldsymbol{H}_N$.

Lemma 6.3. *[35, 118, 148] Let* $h[k] = \frac{\sin(2\pi Wk)}{\pi k}$ *with* $W = \frac{1}{4}$. *Fix* $\epsilon \in (0, \frac{1}{2})$. *Then there exist constants* C_1, C_2 *and* N_1 *such that the distance between any 2 consecutive eigenvalues of* $\boldsymbol{H}_N$ *inside* $(\epsilon, 1-\epsilon)$ *is bounded from below by* $\frac{C_1}{\ln(N)}$ *and from above by* $\frac{C_2}{\ln(N)}$*; that is*

$$\frac{C_1}{\ln(N)} \leq \lambda_l(\boldsymbol{H}_N) - \lambda_{l+1}(\boldsymbol{H}_N) \leq \frac{C_2}{\ln(N)}$$

for all $\epsilon \leq \lambda_{l+1}(\boldsymbol{H}_N) \leq \lambda_l(\boldsymbol{H}_N) \leq 1-\epsilon$ *and* $N \geq N_1$. *Also*

$$\lambda_{\lfloor \frac{1}{2}N \rfloor - 1} \geq \frac{1}{2} \geq \lambda_{\lceil \frac{1}{2}N \rceil}$$

for all $N \in \mathbb{N}$.

On the other hand, we have the following result on the eigenvalues of $\overline{\boldsymbol{C}}_N$. Its proof is given in Appendix D.8.

Lemma 6.4. *Let* $h[k] = \frac{\sin(2\pi Wk)}{\pi k}$ *with* $W = \frac{1}{4}$. *Then*

$$\left| \lambda_l\left(\overline{\boldsymbol{C}}_N\right) - \frac{1}{2} \right| \begin{cases} = 0, & l = N/4, 3N/4, \\ \geq \alpha, & l \in [N] \text{ and } l \neq N/4, 3N/4, \end{cases}$$

with $\alpha = 0.4$ *if* N *is a multiple of* 4.

With more sophisticated analysis, we believe that the above result could be improved to $\alpha \approx 0.45$. This is suggested by Figure 6.3(e).

Combining Lemmas 6.3 and 6.4, we conclude that $\max_{l \in [N]} \left| \lambda_l(\boldsymbol{H}_N) - \lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N) \right|$ approaches ≈ 0.2 as $N \to \infty$ and N is a multiple of 4.

Finally, Figure 6.3(f) plots $\left| \lambda_0(\boldsymbol{H}_N) - \lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N) \right|$ and $\left| \lambda_{N-1}(\boldsymbol{H}_N) - \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N) \right|$ against the dimension N. As can be observed, the largest and smallest eigenvalues of $\overline{\boldsymbol{C}}_N$ converge to the largest and smallest eigenvalues of $\boldsymbol{H}_N$, respectively. This is as guaranteed by Theorem 6.5.

CHAPTER 7

TIME-FREQUENCY LIMITING OPERATORS ON GROUPS

One of the important pieces of progress in harmonic analysis made in last century is the definition of the Fourier transform on locally compact abelian groups (i.e., harmonic analysis on groups) [106]. This framework for harmonic analysis on groups not only unifies the CTFT, DTFT, and DFT (for signal domains, or groups, corresponding to $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{Z}_N := \{0, 1, \ldots, N-1\}$, respectively), but it also allows these transforms to be generalized to other signal domains. This, in turn, makes possible the analysis developed in Chapters 3-5 for other parameterized subspace models appearing in applications such as steerable principal component analysis (PCA) [133] where the domain is the rotation angle on $[0, 2\pi)$, an imaging system with a pupil of finite size [38], line-of-sight (LOS) communication systems with orbital angular momentum (OAM)-based orthogonal multiplexing techniques [136], and many other applications such as those involving rotations in three dimensions [25, Chapter 5].

In this chapter[14], we review existing results on the eigenvalues of composite time- and band-limiting operators and generalize these results to locally compact abelian groups[15]. Applications of this unifying treatment are discussed in relation to channel capacity and to representation and approximation of signals obeying certain parameterized subspace models.

7.1 The Effective Dimensionality of a Signal Family

One of the useful applications of characterizing the spectrum of Toeplitz operators (see (7.4) and the following sentence for the definition of Toeplitz operators) is in computing the effective dimensionality (or the number of degrees of freedom) of signals obeying certain parameterized subspace models. In this section, we formalize this notion of effective dimensionality for a set of functions defined on a group $\mathbb{G}$.

[14]This work is in collaboration with Michael B. Wakin.

[15]The formal definitions of groups and Fourier transform for functions defined on groups are in Section 2.10.

7.1.1 Definitions

To begin, suppose $\mathbb{A}$ is a subset of $\mathbb{G}$ and let $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)) \subset L_2(\mathbb{A})$ denote the set of functions controlled by the symbol $\widehat{\phi}(\xi)$:

$$\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)) := \left\{ x \in L_2(\mathbb{A}) : x(g) = \int_{\widehat{\mathbb{G}}} \alpha(\xi)\widehat{\phi}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi, \int |\alpha(\xi)|^2\,\mathrm{d}\,\xi \leq 1, g \in \mathbb{A} \right\}, \quad (7.1)$$

which is a subset of $L_2(\mathbb{A})$. We note that in (7.1), the symbol $\widehat{\phi}(\xi)$ is fixed and we will discuss its role soon. It is clear that x obeys a parameterized subspace model as described in Section 1.2.

Also let $\mathbb{M}_n \subset L_2(\mathbb{G})$ denote an n-dimensional subspace of $L_2(\mathbb{G})$. The distance between a point $x \in L_2(\mathbb{G})$ and the subspace $\mathbb{M}_n$ is defined as

$$d(x, \mathbb{M}_n) := \inf_{y \in \mathbb{M}_n} \int (x(g) - y(g))^2\,\mathrm{d}\,g = \int (x(g) - (\mathcal{P}_{\mathbb{M}_n} x)(g))^2\,\mathrm{d}\,g = \sup_{\substack{z \in L_2(\mathbb{G}) \\ z \perp \mathbb{M}_n}} \frac{\left|\langle x, z\rangle_{L_2(\mathbb{G})}\right|}{\|z\|_{L_2(\mathbb{G})}}, \quad (7.2)$$

where $\mathcal{P}_{\mathbb{M}_n} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ represents the orthogonal projection onto the subspace $\mathbb{M}_n$. We define the distance $d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n)$ between the set $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ and the subspace $\mathbb{M}_n$ as follows:

$$d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n) := \sup_{x \in \mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))} d(x, \mathbb{M}_n) = \sup_{x \in \mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))} \inf_{y \in \mathbb{M}_n} \int (x(g) - y(g))^2\,\mathrm{d}\,g,$$

which represents the largest distance from the elements in $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ to the subspace $\mathbb{M}_n$. The Kolmogorov n-width [76], $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$ of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ in $L_2(\mathbb{G})$ is defined as the smallest distance $d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n)$ over all n-dimensional subspaces of $L_2(\mathbb{G})$; that is

$$d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) := \inf_{\mathbb{M}_n} d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n). \quad (7.3)$$

In summary, the n-width $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$ characterizes how well the set $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ can be approximated by an n-dimensional subspace of $L_2(\mathbb{G})$. By its definition, $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$

is non-increasing in terms of the dimensionality n. For any fixed $\epsilon > 0$, we define the *effective dimensionality*, or *number of degrees of freedom*, of the set $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ at level ϵ as [50]

$$\mathcal{N}(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \epsilon) = \min \left\{ n : d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) < \epsilon \right\}.$$

In words, the above definition ensures that there exists a subspace $\mathbb{M}_n$ of dimensionality $n = \mathcal{N}(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \epsilon)$ such that for every function $x \in \mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$, one can find at least one function $y \in \mathbb{M}_n$ so that the distance between x and y is at most ϵ.

We note that the reason we impose an energy constraint on the elements x of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ in (7.1) is that we use the absolute distance to quantify the proximity of x to the subspace $\mathbb{M}_n$ in (7.2).

7.1.2 Connection to Operators

In order to compute $\mathcal{N}(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \epsilon)$, we may define an operator $\mathcal{A} : L_2(\widehat{\mathbb{G}}) \to L_2(\mathbb{A})$ as

$$(\mathcal{A}\alpha)(g) = \int_{\widehat{\mathbb{G}}} \alpha(\xi)\widehat{\phi}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi, g \in \mathbb{A}.$$

The adjoint $\mathcal{A}^* : L_2(\mathbb{A}) \to L_2(\widehat{\mathbb{G}})$ is given by

$$(\mathcal{A}^* x)(\xi) = \int_{\mathbb{A}} x(g)\widehat{\phi^*}(\xi)\chi_\xi^*(g)\,\mathrm{d}\,g.$$

The composition of $\mathcal{A}$ and $\mathcal{A}^*$ gives a self-adjoint operator $\mathcal{A}\mathcal{A}^* : L_2(\mathbb{A}) \to L_2(\mathbb{A})$ as follows:

$$\begin{aligned}(\mathcal{A}\mathcal{A}^* x)(g) &= \int_{\widehat{\mathbb{G}}} \widehat{\phi}(\xi)\chi_\xi(g) \int_{\mathbb{A}} x(h)\widehat{\phi^*}(\xi)\chi_\xi^*(h)\,\mathrm{d}\,h\,\mathrm{d}\,\xi \\ &= \int_{\mathbb{A}} x(h) \int_{\widehat{\mathbb{G}}} \left|\widehat{\phi}(\xi)\right|^2 \chi_\xi(h^{-1} \circ g)\,\mathrm{d}\,\xi\,\mathrm{d}\,h \\ &= \int_{\mathbb{A}} x(h)(\phi \star \phi^*)(h^{-1} \circ g)\,\mathrm{d}\,h,\end{aligned} \tag{7.4}$$

where $\phi(g) = \int_{\widehat{\mathbb{G}}} \widehat{\phi}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi$ is the inverse Fourier transform of $\widehat{\phi}$. Because this linear operator involves a kernel function $\phi \star \phi^*(h^{-1} \circ g)$ that depends only on the difference $^{-1} \circ g$, we refer to it as a *Toeplitz operator*.[16]

[16] Our notion of Toeplitz operators follows from the definition of Toeplitz operators in [56, Section 7.2].

The following result will help in computing $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$ and the effective dimensionality of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ and choosing the optimal basis for representing the elements of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$.

Proposition 7.1. *Let the eigenvalues of $\mathcal{A}\mathcal{A}^*$ be denoted and arranged as $\lambda_1 \geq \lambda_2 \geq \cdots$. Then the n-width of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ can be computed as*

$$d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) = \sqrt{\lambda_n},$$

and the optimal n-dimensional subspace to represent $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ is the subspace spanned by the first n eigenvectors of $\mathcal{A}\mathcal{A}^*$.

The proof of Proposition 7.1 is given in Appendix E.1.

7.2 Time-Frequency Limiting Operators on Locally Compact Abelian Groups

Now we consider the time-frequency limiting operators on locally compact abelian groups. As we have briefly explained before, time-frequency limiting operators in the context of the classical groups where $\mathbb{G}$ are the real-line, $\mathbb{Z}$, and $\mathbb{Z}_N$ play important roles in signal processing and communication. By considering time-frequency limiting operators on locally compact abelian groups, we aim to (i) provide a unified treatment of the previous results on the eigenvalues of the operators resulting in PSWF's, DPSS's, and periodic DPSS's (PDPSS's) [71, 57]; and (ii) extend these results to other signal domains such as rotations in a plane and three dimensions [25, Chapter 5]. In particular, we will investigate the eigenvalues of time-frequency limiting operators on locally compact abelian groups and show that they exhibit similar behavior to both the conventional continuous time and discrete time settings: when sorted by magnitude, there is a cluster of eigenvalues close to (but not exceeding) 1, followed by an abrupt transition, after which the remaining eigenvalues are close to 0. This behavior also resembles the rectangular shape of the frequency response of the original band-limiting operator. We will also discuss the applications of this unifying treatment in relation to channel capacity and to representation and approximation of signals.

To introduce the time-frequency limiting operators, consider two subsets $\mathbb{A} \in \mathbb{G}$ and $\mathbb{B} \in \widehat{\mathbb{G}}$. Define $\mathcal{T}_{\mathbb{A}} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ as a time-limiting operator that makes a function zero

outside $\mathbb{A}$. Also define $\mathcal{B}_{\mathbb{B}} = \mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ as a band-limiting operator that takes the Fourier transform of an input function on $L_2(\mathbb{G})$, sets it to zero outside $\mathbb{B}$, and then computes the inverse Fourier transform. The operator $\mathcal{B}_{\mathbb{B}}$ acts on $L_2(\mathbb{G})$ as a convolutional integral operator:

$$\begin{aligned}(\mathcal{B}_{\mathbb{B}}x)(g) &= \int_{\mathbb{B}} \widehat{x}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi \\ &= \int_{\mathbb{B}} \left(\int_{\mathbb{G}} x(h)\chi_\xi^*(h)\,\mathrm{d}\,h \right) \chi_\xi(g)\,\mathrm{d}\,\xi \\ &= \int_{\mathbb{G}} K_{\mathbb{B}}(h^{-1}\circ g)x(h)\,\mathrm{d}\,h,\end{aligned}$$

where

$$K_{\mathbb{B}}(h^{-1}\circ g) = \int_{\mathbb{B}} \chi_\xi^*(h)\chi_\xi(g)\,\mathrm{d}\,\xi = \int_{\mathbb{B}} \chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi. \tag{7.5}$$

It is of interest to study the eigenvalues of the following operators

$$\mathcal{O}_{\mathbb{A},\mathbb{B}} = \mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}, \quad \text{and} \quad \mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}. \tag{7.6}$$

Utilizing the expression for $\mathcal{B}_{\mathbb{B}}$, the operator $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ acts on any $x \in L_2(\mathbb{G})$ as follows

$$\left(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}x\right)(g) = \begin{cases} \int_{\mathbb{A}} K_{\mathbb{B}}(h^{-1}\circ g)x(h)\,\mathrm{d}\,h, & g \in \mathbb{A} \\ 0, & \text{otherwise.}\end{cases}$$

The operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ is symmetric and completely continuous and we denote its eigenvalues by $\lambda_\ell(\mathcal{O}_{\mathbb{A},\mathbb{B}})$. Due to the time- and band-limiting characteristics of the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$, the eigenvalues of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ are between 0 and 1. To see this, let $x(g) \in L_2(\mathbb{A})$:

$$\begin{aligned}\langle (\mathcal{O}_{\mathbb{A},\mathbb{B}}x)(g), x(g)\rangle &= \left\langle \int_{\mathbb{A}}\int_{\mathbb{B}} \chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi x(h)\,\mathrm{d}\,h, x(g) \right\rangle \\ &= \int_{\mathbb{B}} \left(\int_{\mathbb{A}}\int_{\mathbb{A}} \chi_\xi(h^{-1}\circ g)x(h)x^*(g)\,\mathrm{d}\,h\,\mathrm{d}\,g \right) \mathrm{d}\,\xi \\ &= \int_{\mathbb{B}} |\widehat{x}(\xi)|^2\,\mathrm{d}\,\xi \geq 0.\end{aligned}$$

On the other hand, we have

$$\int_{\mathbb{B}} |\widehat{x}(\xi)|^2 \,\mathrm{d}\,\xi \leq \int_{\widehat{\mathbb{G}}} |\widehat{x}(\xi)|^2 \,\mathrm{d}\,\xi = \int_{\mathbb{G}} |x(g)|^2 \,\mathrm{d}\,g.$$

7.2.1 Eigenvalue Distribution of Time-Frequency Limiting Operators

To investigate the eigenvalues of the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}} = \mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$, we first note that without the time-limiting operator $\mathcal{T}_{\mathbb{A}}$, the eigenvalues of $\mathcal{B}_{\mathbb{B}}$ are simply given by the Fourier transform of $K_{\mathbb{B}}(g)$, and thus they are either 1 or 0. Our main question is how the spectrum of the time-frequency limiting operator relates to the spectrum of the band-limited operator. Based on the binary spectrum of $\mathcal{B}_{\mathbb{B}}$, we expect that the eigenvalues of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ to have a particular behavior: when sorted by magnitude, there should be a cluster of eigenvalues close to (but not exceeding) 1, followed by an abrupt transition, after which the remaining eigenvalues should be close to 0. Moreover, the number of effective (i.e., relatively large) eigenvalues should be essentially equal to the time-frequency area $|\mathbb{A}||\mathbb{B}|$. These results are confirmed below and reveal the dimensionality (or the number of degrees of freedom) of classes of band-limited signals observed over a finite time, which is fundamental to characterizing the performance limits of communication systems.

Before presenting the main results, we introduce new notation for subsets of $\mathbb{G}$ (or $\widehat{\mathbb{G}}$) which are asymptotically increasing to cover the whole group. This is similar to how we discussed the case where $N \to \infty$ in Section 2.8. To that end, let $\mathbb{A}_\tau, \tau \in (0,\infty)$ be a system of open subsets of $\mathbb{G}$ such that $0 < \mu(\mathbb{A}_\tau) < \infty$. The subscript τ is sometimes dropped when it is clear from the context. We can view $\mathbb{A}_\tau$ as a set of open subsets that depend on the parameter τ. One can also define a system of open subsets with multiple parameters. We now preset one of our main results concerning the asymptotic behavior for the eigenvalues of the time-frequency limiting operators $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$ when $\mathbb{A}_\tau$ approaches $\mathbb{G}$.

Theorem 7.1. *Suppose $\mathbb{B}$ is a fixed subset of $\widehat{\mathbb{G}}$ and let $\epsilon \in (0, \frac{1}{2})$. Let*

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; (a,b)) := \#\left\{\ell : a < \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) < b\right\}$$

denote the number of eigenvalues of $\mathcal{T}_{\mathbb{A}_\tau}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}_\tau}$ that is between a and b. Then if

$$\lim_{\tau \to \infty} \mathbb{A}_\tau = \mathbb{G} \tag{7.7}$$

holds almost everywhere, we have

$$\sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{A}_\tau||\mathbb{B}|, \tag{7.8}$$

$$\sum_{\ell} \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{A}_\tau||\mathbb{B}| - o(|\mathbb{A}_\tau||\mathbb{B}|), \tag{7.9}$$

which immediately implies

$$\lim_{\tau \to \infty} \frac{\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; [1-\epsilon, 1])}{|\mathbb{A}_\tau|} = |\mathbb{B}|, \tag{7.10}$$

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; (\epsilon, 1-\epsilon)) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{7.11}$$

The proof of Theorem 7.1 is given in Appendix E.2. Theorem 7.1 formally confirms that the spectra of the time-frequency limiting operators resemble the rectangular shape of the spectrum of the band-limiting operator. As guaranteed by (7.10), the number of effective eigenvalues of the time-frequency limiting operator is asymptotically equal to the time-frequency area $|\mathbb{A}_\tau||\mathbb{B}|$. Similar results for time-frequency limiting operators in the context of classical groups where $\mathbb{G}$ is $\mathbb{R}^n$ are given in [79, 51]. We discuss the applications of Theorem 7.1 in channel capacity and representation and approximation of signals in more detail in the following subsections.

As we explained before, the time-frequency limiting operators in the context of the classical groups where $\mathbb{G}$ are the real-line, $\mathbb{Z}$, and $\mathbb{Z}_N$ were first studied by Landau, Pollak, and Slepian who wrote a series of papers regarding the dimensionality of time-limited signals that are approximately band-limited (or vice versa) [120, 81, 82, 115, 118] (see also [117, 119] for concise overviews of this body of work). After that, a set of results concerning the number of eigenvalues within the transition region $(0, 1)$ have been established

in [83, 45, 101, 69, 75, 143]. which will be reviewed in detail in the following remarks.

Remark 7.1. Using the explicit expressions for the character function $\chi_\xi(g)$ and the kernel $K_{\mathbb{B}}(g)$ and applying integration by parts for (E.2), one can improve the second term in (7.9) to $O(\log(|\mathbb{A}_\tau||\mathbb{B}|))$ for many common one-dimensional cases:

- Suppose $\mathbb{G} = \mathbb{R}$ and $\widehat{\mathbb{G}} = \mathbb{R}$. Let $\mathbb{A}_T = [-\frac{T}{2}, \frac{T}{2}]$ (where $\tau = T$ in (7.7)) and $\mathbb{B} = [-\frac{1}{2}, \frac{1}{2}]$ without loss of generality. Then the kernel $K_{\mathbb{B}}(t)$ turns out to be

$$K_{\mathbb{B}}(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi Ft} \,\mathrm{d}\, F = \frac{2\sin(\pi t)}{\pi t}.$$

  Plugging in this form into (E.2) gives [63]

$$\sum_\ell (\lambda_\ell(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}))^2 = T - O(\log(T)).$$

  In this case, the eigenfunctions of the operator $\mathcal{O}_{\mathbb{A}_T,\boldsymbol{B}}$ are known as PSWF's.

- As an another example, suppose $\mathbb{G} = \mathbb{Z}$, $\widehat{\mathbb{G}} = [-\frac{1}{2}, \frac{1}{2}]$ and let $\mathbb{A}_N = \{0, 1 \ldots, N-1\}$ (where $\tau = N$ in (7.7)), $\mathbb{B} = [-W, W]$ with $W \in (0, \frac{1}{2})$. In this case, the kernel $K_{\mathbb{B}}(t)$ becomes

$$K_{\mathbb{B}}[n] = \int_{-W}^{W} e^{j2\pi fn} \,\mathrm{d}\, f = \frac{\sin(2\pi Wn)}{\pi Wn}.$$

  Then plugging in this form into (E.2) gives (see also Theorem 5.1)

$$\sum_\ell (\lambda_\ell(\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}))^2 = 2NW - O(\log(2NW)).$$

  We note that in this case, the operator $\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}$ is equivalent to the $N \times N$ prolate matrix $\boldsymbol{B}_{N,W}$ with entries

$$\boldsymbol{B}_{N,W}[m,n] := \frac{\sin\left(2\pi W(m-n)\right)}{\pi(m-n)} \tag{7.12}$$

  for all $m, n \in \{0, 1, \ldots, N-1\}$. The eigenvalues and eigenvectors of the matrix $\boldsymbol{B}_{N,W}$ are referred to as the DPSS eigenvalues and DPSS vectors, respectively.

- As a final example, we consider $\mathbb{G} = \mathbb{Z}_N$, $\widehat{\mathbb{G}} = \mathbb{Z}_N$ and the Fourier transform is the conventional DFT. Suppose $M, K \leq N$. Let

$$\mathbb{A}_M = \{0, 1, \ldots, M-1\}, \ \mathbb{B} = \{0, 1, \ldots, K-1\}, \tag{7.13}$$

where $\tau = M$ in (7.7). In this case, $\chi_k[n] = e^{j2\pi \frac{nk}{N}}$ and the kernel $K_{\mathbb{B}}[n]$ is

$$K_{\mathbb{B}}[n] = \sum_{k=0}^{K-1} e^{j2\pi \frac{nk}{N}} = e^{j\pi n \frac{K-1}{N}} \frac{\sin(\pi \frac{nK}{N})}{\sin(\pi \frac{n}{N})}.$$

Then plugging in this form into (E.2) gives [45, 143]

$$\sum_l (\lambda_\ell(\mathcal{O}_{\mathbb{A}_M,\mathbb{B}}))^2 = \frac{MK}{N} - O(\log(\frac{MK}{N})).$$

Through the above examples, one may wonder whether we can in general replace the second term in (7.9) by $O(\log(|\mathbb{A}_\tau||\mathbb{B}|))$ with a finer analysis of $\sum_l (\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}))^2$. We utilize a two-dimensional example to answer this question in the negative:

- Suppose $\mathbb{G} = \mathbb{Z}^2$, $\widehat{\mathbb{G}} = [-\frac{1}{2}, \frac{1}{2}]$ and let $\mathbb{A} = \{0, 1 \ldots, N-1\} \times \{0, 1 \ldots, N-1\}, \mathbb{B} = [-W, W] \times [-W, W]$ with $W \in (0, \frac{1}{2})$. In this case, the kernel $K_{\mathbb{B}}[n_1, n_2]$ is

$$K_{\mathbb{B}}[n_1, n_2] = \int_{-W}^{W} \int_{-W}^{W} e^{j2\pi f_1 n_1} e^{j2\pi f_2 n_2} \, \mathrm{d}\, f_1 \, \mathrm{d}\, f_2 = \frac{\sin(2\pi W n_1)}{\pi W n_1} \frac{\sin(2\pi W n_2)}{\pi W n_2}.$$

The eigenvectors of the corresponding operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ are known as the two-dimensional DPSS's. For this case, we have

$$\sum_\ell (\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}))^2 = 4N^2W^2 - O(NW \log(NW)).$$

In other words, we can only improve the second term in (7.9) to $O(NW \log(NW))$ rather than $O(\log(4N^2W^2))$.

Remark 7.2. We note that the transition region in (7.11) depends on ϵ in the form of $O(\frac{1}{\epsilon(1-\epsilon)})$ because it is simply derived from (7.8) and (7.9). A better understanding of the transition region requires further complicated analysis. In the literature, finer results on the transition region are known for several common cases:

- The results for the eigenvalue distribution of the continuous time-frequency localization operator (where $\mathbb{G} = \mathbb{R}$, $\widehat{\mathbb{G}} = \mathbb{R}$, $\mathbb{A}_T = [-\frac{T}{2}, \frac{T}{2}]$ and $\mathbb{B} = [-\frac{1}{2}, \frac{1}{2}]$) has a rich history. As one example, for any $\epsilon \in (0, 1)$, Landau and Widom [83] provided the following asymptotic result on $\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; [\epsilon, 1])$:

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; [\epsilon, 1]) = T + \left(\frac{1}{\pi^2} \log \frac{1-\epsilon}{\epsilon}\right) \log \frac{\pi T}{2} + o\left(\log \frac{\pi T}{2}\right).$$

This asymptotic result ensures the $O(\log(\frac{1}{\epsilon}) \log(T))$ dependence on ϵ and time-frequency area T. Recently, Osipov [101] proved that

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; [\epsilon, 1]) \leq T + C \log(T)^2 \log(1/\epsilon),$$

where C is a constant. Israel [69] provided a non-asymptotic bound on the number of eigenvalues in the transition region. Fix $\eta \in (0, 1/2]$. Given $\epsilon \in (0, 1/2)$ and $T \geq 2$, then [69]

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; (\epsilon, 1-\epsilon)) \leq 2C_\eta \left(\log\left(\frac{\log T}{\epsilon}\right)\right)^{1+\eta} \log\left(\frac{T}{\epsilon}\right), \tag{7.14}$$

where C_η is a constant dependent on $\eta \in (0, \frac{1}{2}]$.

- The earliest result on the eigenvalue distribution of the discrete time-frequency localization operator (where $\mathbb{G} = \mathbb{Z}$, $\widehat{\mathbb{G}} = [-\frac{1}{2}, \frac{1}{2})$, $\mathbb{A}_N = \{0, 1, \ldots, N-1\}$ and $\mathbb{B} = [-W, W]$ with $W \in (0, \frac{1}{2})$) comes from Slepian [118], where it is shown that for any $b \in \mathbb{R}$, asymptotically the DPSS eigenvalue $\lambda_\ell(\mathcal{O}(\mathbb{A}_N, \mathbb{B})) \to \frac{1}{1+e^{\pi b}}$ as $N \to \infty$ if $\ell = \lfloor 2NW + \frac{b}{\pi} \log N \rfloor$. This implies the asymptotic result:

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}; (\epsilon, 1-\epsilon)) \sim \frac{2}{\pi^2} \log N \log\left(\frac{1}{\epsilon} - 1\right).$$

Recently, by examining the difference between the operator $\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}$ and the one formed by a partial DFT matrix, we have shown in Corollary 3.1 the following nonasymptotic

result characterizing the $O(\log N \log \frac{1}{\epsilon})$ dependence:

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}; (\epsilon, 1-\epsilon)) \leq \left(\frac{8}{\pi^2}\log(8N) + 12\right)\log\left(\frac{15}{\epsilon}\right). \tag{7.15}$$

Theorem 3.2 also provides similar results for the eigenvalue distribution of discrete periodic time-frequency localization operator with sets $\mathbb{A}_M$ and $\mathbb{B}$ defined in (7.13):

$$\begin{aligned}&\mathcal{N}(\mathcal{O}_{\mathbb{A}_M,\mathbb{B}}; (\epsilon, 1-\epsilon))\\ &\leq \left(\frac{8}{\pi^2}\log(8N) + 12\right)\log\left(\frac{15}{\epsilon}\right) + 4\max\left(\frac{-\log\left(\frac{\pi}{32}\left(\left(\frac{M}{N}\right)^2 - 1\right)\epsilon\right)}{\log\left(\frac{M}{N}\right)}, 0\right).\end{aligned}$$

Remark 7.3. It is also of particular interest to have a finer result on the number of eigenvalues that is greater than $\frac{1}{2}$ since this together with the size of the transition region gives us a complete understanding of the eigenvalue distribution.

- Landau [80] establishes the number of PSWF eigenvalues that are greater than $\frac{1}{2}$ as follows

$$\lambda_{(\lfloor T \rfloor - 1)}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}) \geq \frac{1}{2} \geq \lambda_{(\lceil T \rceil)}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}). \tag{7.16}$$

- Theorem 5.2 provides a similar result for the DPSS eigenvalues.

In the following two subsections, we review some applications of Theorem 7.1.

7.3 Application: Communication

In [51], Franceschetti extended Landau's theorem for simple time and frequency intervals to other time and frequency sets of complicated shapes. Lim and Franceschetti [89] related the number of degrees of freedom of the space of bandlimited signals to the deterministic notions of capacity and entropy. Now we apply Theorem 7.1 to the effective dimensionality of the bandlimited signals observed over a finite set $\mathbb{A}$ by utilizing the result in Section 7.1.

To that end, we plug $\widehat{\phi}(\xi) = 1_{\mathbb{B}}(\xi) = \begin{cases} 1, & \xi \in \mathbb{B} \\ 0, & \xi \notin \mathbb{B} \end{cases}$ the indicator function on $\mathbb{B}$ into (7.1) and get the following set of bandlimited functions observed only over $\mathbb{A}$:

$$\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi)) := \left\{ x \in L_2(\mathbb{A}) : x(g) = \int_{\mathbb{B}} \alpha(\xi) \chi_{\xi}(g) \, \mathrm{d}\, \xi, \int |\alpha(\xi)|^2 \leq 1, g \in \mathbb{A} \right\}.$$

When $\mathbb{A} \subset \mathbb{R}^2$ represents a subset of time and space, the number of degrees of freedom in the set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$ determines the total amount of information can be transmitted in time and space by multiple-scattered electromagnetic waves [51]. Now we turn to compute the effective dimensionality of the general set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$. In this case, $|\widehat{\phi}(\xi)|^2 = 1_{\mathbb{B}}(\xi)$ and the corresponding operator $\mathcal{A}\mathcal{A}^*$ defined in (7.4) is equivalent to the time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ in (7.6). Now Proposition 7.1 implies that the effective dimensionality $\mathcal{N}(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \epsilon)$ is equal to the number of eigenvalues of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ that is greater than ϵ, which is given by Theorem 7.1. In words, the effective dimensionality of the set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$ is essentially $|\mathbb{A}||\mathbb{B}|$, and is insensitive to the level ϵ (as illustrated before, in many cases, this dimensionality only has $\log(\frac{1}{\epsilon})$ dependence on ϵ).

7.4 Application: Signal Representation

In addition to the eigenvalues of the time-frequency limiting operator $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$, the eigenfunctions $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ are also of significant importance, owing to their concentration in the time and frequency domains. To see this, let $u_\ell(g)$ be the ℓ-th eigenfunction of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$, corresponding to the ℓ-th eigenvalue $\lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}})$. Denoting the Fourier transform of $u_\ell(g)$ by $\widehat{u}_\ell(\xi)$, we have

$$\begin{aligned} \int_{\mathbb{B}} |\widehat{u}_\ell(\xi)|^2 \, \mathrm{d}\, \xi &= \langle \mathcal{T}_{\mathbb{B}} \mathcal{F} u_\ell, \mathcal{T}_{\mathbb{B}} \mathcal{F} u_\ell \rangle \\ &= \left\langle \mathcal{F}^{-1} \mathcal{T}_{\mathbb{B}} \mathcal{F} u_\ell, u_\ell \right\rangle \\ &= \left\langle \mathcal{F}^{-1} \mathcal{T}_{\mathbb{B}} \mathcal{F} \mathcal{T}_{\mathbb{A}} u_\ell, \mathcal{T}_{\mathbb{A}} u_\ell \right\rangle \\ &= \left\langle \mathcal{T}_{\mathbb{A}} \mathcal{F}^{-1} \mathcal{T}_{\mathbb{B}} \mathcal{F} \mathcal{T}_{\mathbb{A}} u_\ell, \mathcal{T}_{\mathbb{A}} u_\ell \right\rangle \\ &= \lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}) \|\mathcal{T}_{\mathbb{A}} u_\ell\|^2, \end{aligned} \tag{7.17}$$

where the third line follows because $u_\ell(g)$ is a bandlimited signal (i.e., $\mathcal{T}_{\mathbb{A}}(u_\ell) = u_\ell$), and the last line utilizes $\mathcal{T}_{\mathbb{A}}\mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F}\mathcal{T}_{\mathbb{A}}u_\ell = \mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}u_\ell = \lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}})u_\ell$. In words, (7.17) states that the eigenfunctions u_ℓ have $\lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{A}_{\mathbb{A}})$ proportion of energies within the band $\mathbb{B}$, implying that though the eigenfunctions are not exactly bandlimited, their Fourier transform is mostly concentrated in the band $\mathbb{B}$ when $\lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{A}_{\mathbb{A}})$ is close to 1. Thus, the first $\approx |\mathbb{A}||\mathbb{B}|$ eigenfunctions can be utilized as window functions for spectral estimation, and as a highly efficient basis for representing bandlimited signals that are observed over a finite set $\mathbb{A}$.

Recall that $\mathbb{W}(\mathbb{A}_\tau, 1_{\mathbb{B}}(\xi))$ (defined in (7.1)) consists of bandlimited signals observed over a finite set $\mathbb{A}_\tau$. Applying Proposition 7.1, we compute the n-width of the set $\mathbb{W}(\mathbb{A}_\tau, 1_{\mathbb{B}}(\xi))$ as follows:

$$d_n(\mathbb{W}(\mathbb{A}_\tau, 1_{\mathbb{B}}(\xi))) = \sqrt{\lambda_n(\mathcal{A}\mathcal{A}^*)} = \sqrt{\lambda_n(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})}.$$

By the definition of (7.3), we know for any $x(g) \in \mathbb{W}(\mathbb{A}_\tau, 1_{\mathbb{B}}(\xi))$,

$$\int_{\mathbb{A}_\tau} |x(g) - (\mathcal{P}_{\mathbb{U}_n}x)(g)|^2 \,\mathrm{d}\, g \leq \sqrt{\lambda_n(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})},$$

where $\mathbb{U}_n$ is the subspace spanned by the first n eigenvectors of $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$, i.e.,

$$\mathbb{U}_n := \mathrm{span}\{u_0(g), u_1(g), \ldots, u_{n-1}(g)\}. \tag{7.18}$$

Now we utilize Theorem 7.1 to conclude that the representation residual $\sqrt{\lambda_n(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})}$ is very small by choosing $n \approx |\mathbb{A}_\tau||\mathbb{B}|$.

We now investigate the basis $\mathbb{U}_n$ for representing time-limited version of characters $\chi_\xi(g)$ and bandlimited signals.

7.4.1 Approximation Quality for Time-Limited Version of Characters $\chi_\xi(g)$

We first restrict our focus to the simplest possible bandlimited signals that are observed over a finite period: pure characters $\chi_\xi(g)$ when g is limited to $\mathbb{A}_\tau$. Without knowing the exact carrier frequency ξ in advance, we attempt to find an efficient low-dimensional basis for capturing the energy in any signal $\chi_\xi(g)$. To that end, we let $\mathbb{M}_n \in L_2(\mathbb{A}_\tau)$ denote an

n-dimensional subspace of $L_2(\mathbb{A}_\tau)$. We would like to minimize

$$\int_{\mathbb{B}} \|\chi_\xi - \mathcal{P}_{\mathbb{M}_n}\chi_\xi\|^2_{L_2(\mathbb{A}_\tau)}\,\mathrm{d}\,\xi. \tag{7.19}$$

The following result establishes that the degree of approximation accuracy in a MSE sense provided by the subspace $\mathbb{U}_n$ for representing the time-limited version of characters $\chi_\xi(g)$ (where g is limited to $\mathbb{A}_\tau$).

Theorem 7.2. *For any $n \in \mathbb{Z}^+$, the optimal n-dimensional subspace which minimizes* (7.19) *is $\mathbb{U}_n$. Furthermore, with this choice of subspace, we have*

$$\frac{1}{|\mathbb{B}|}\int_{\mathbb{B}} \frac{\|\chi_\xi - \mathcal{P}_{\mathbb{U}_n}\chi_\xi\|^2_{L_2(\mathbb{A}_\tau)}}{\|\chi_\xi\|^2_{L_2(\mathbb{A}_\tau)}}\,\mathrm{d}\,\xi = 1 - \frac{\sum_{\ell=0}^{n-1}\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})}{|\mathbb{A}_\tau||\mathbb{B}|}.$$

The proof of Theorem 7.2 is given in Appendix E.3. Combined with Theorem 7.1, Theorem 7.2 implies that by choosing $n \approx |\mathbb{A}_\tau||\mathbb{B}|$, in average the subspace spanned by the first n eigenfunctions of $\mathcal{T}_{\mathbb{A}_\tau}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}_\tau}$ is expected to accurately represent time-limited characters within the band of interest. We note that the representation guarantee for time-limited characters $\{\mathcal{T}_{\mathbb{A}_\tau}\chi_\xi, \xi \in \mathbb{B}\}$ can also be used for most bandlimited signals that are observed over a finite set $\mathbb{A}_\tau$. To see this, suppose $x(g)$ is a bandlimited function which can be represented as

$$x(g) = \int_{\mathbb{B}} \widehat{x}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi.$$

An immediate consequence of the above equation is to view $\{\mathcal{T}_{\mathbb{A}_\tau}\chi_\xi, \xi \in \mathbb{B}\}$ as the atoms for building $\mathcal{T}_{\mathbb{A}_\tau}x$:

$$\mathcal{T}_{\mathbb{A}_\tau}x = \int_{\mathbb{B}} \widehat{x}(\xi)\mathcal{T}_{\mathbb{A}_\tau}\chi_\xi\,\mathrm{d}\,\xi.$$

7.4.2 Approximation Quality for Random Bandlimited Signals

We can also approach the representation ability of the subspace $\mathbb{U}_n$ (defined in (7.18)) from a probabilistic perspective.

Theorem 7.3. *Let $x(g) = \chi_\xi(g), g \in \mathbb{A}_\tau$ be a random function where ξ is a random variable with uniform distribution on $\mathbb{B}$. Then we have*

$$\frac{\mathrm{E}\left[\|x - \mathcal{P}_{\mathbb{U}_n} x\|^2_{L_2(\mathbb{A}_\tau)}\right]}{\mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A}_\tau)}\right]} = 1 - \frac{\sum_{\ell=0}^{n-1} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})}{|\mathbb{A}_\tau||\mathbb{B}|}.$$

The proof of Theorem 7.3 is given in Appendix E.4. With this result, we are able to show that in a certain sense, most bandlimited signals, when time-limited, are well-approximated by a signal within the subspace $\mathbb{U}_n$. In particular, the following result establishes than the bandlimited random processes, when time-limited, are in expectation well-approximated.

Corollary 7.1. *Let $x(g), g \in \mathbb{G}$ be a zero-mean wide sense stationary random process over the group $\mathbb{G}$ with power spectrum*

$$P_x(\xi) = \begin{cases} \frac{1}{|\mathbb{B}|}, & \xi \in \mathbb{B}, \\ 0, & \textit{otherwise.} \end{cases}$$

Suppose we only observe x over the set $\mathbb{A}_\tau$. Then we have

$$\frac{\mathrm{E}\left[\|x - \mathcal{P}_{\mathbb{U}_n} x\|^2_{L_2(\mathbb{A}_\tau)}\right]}{\mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A}_\tau)}\right]} = 1 - \frac{\sum_{\ell=0}^{n-1} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})}{|\mathbb{A}_\tau||\mathbb{B}|}$$

The proof of Corollary (7.1) is given in Appendix E.5. As in our discussion following Theorem 7.1, the term $1 - \frac{\sum_{\ell=0}^{n-1} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})}{|\mathbb{A}_\tau||\mathbb{B}|}$ appearing in Theorems 7.3 and Corollary 7.1 can be very small when we choose n slightly larger than $|\mathbb{A}_\tau||\mathbb{B}|$. This suggests that in a probabilistic sense, most bandlimited functions, when time-limited, will be well-approximated by a small number of the eigenfunctions of the operator $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$.

7.5 Applications in the Common Time and Frequency Domains

We now review several applications (that are not included in Chapters 3 and 4) involving the time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ in the common time and frequency domains, where the eigenfunctions correspond to DPSS's PSWF's, and PDPSS's.

It follows from (7.17) that, among all the functions that are time-limited to the set $\mathbb{A}$, the first eigenfuction $u_0(g)$ is maximally concentrated in the set $\mathbb{B}$ of the frequency domain.

Motivated by this result, the first DPSS vector is utilized as a filter for super-resolution [47]. In [125], the first $\approx 2NW$ DPSS vectors are utilized as window functions (a.k.a. tapers) for spectral estimation. The multitaper method [125] averages the tapered estimates with the DPSS vectors, and has been used in a variety of scientific applications including statistical signal analysis [27], geophysics and cosmology [28].

By exploiting the concentration behavior of the PSWF's in the time and frequency domains (where $\mathbb{G} = \mathbb{R}$ and $\widehat{\mathbb{G}} = \mathbb{R}$), Xiao et al. [135] utilized the PSWF's to numerically construct quadratures, interpolation and differentiation formulae for bandlimited functions. Gosse [53] constructed a PSWF dictionary consisting of the first few PSWF's for recovering smooth functions from random samples. The connection between time-frequency localization of multiband signals and sampling theory for such signals is investigated in [70]. In [112, 113], the authors also considered a PSWF dictionary for reconstruction of electroencephalography (EEG) signals and time-limited signals that are also nearly bandlimited from nonuniform samples. Chen and Vaidyanathan [23] utilized the PSWF's to represent the clutter subspace (and hence mitigate the clutter), facilitating space-time adaptive processing for multiple-input multiple-output (MIMO) radar system; see also [138, 42].

We finally mention that the eigenvalue concentration behavior in Theorem 7.1 can also be exploited for solving a linear system involving the Toeplitz operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$: $y = \mathcal{O}_{\mathbb{A},\mathbb{B}}x$. Since the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ has a mass of eigenvalues that are very close to 0, the system is often solved by using the rank-K pseudoinverse of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ wheren $K \approx |\mathbb{A}||\mathbb{B}|$. In the case when the Toeplitz operator is the prolate matrix $\boldsymbol{B}_{N,W}$ defined in (7.12), its truncated pseudoinverse is well approximated as the sum of $\boldsymbol{B}^*_{N,W}$ (which is equal to $\boldsymbol{B}_{N,W}$) and a low-rank matrix [95, 75] since most of the eigenvalues of $\boldsymbol{B}^*_{N,W}$ are either very close to 1 or 0. By utilizing the fact that $\boldsymbol{B}_{N,W}$ is a Toeplitz matrix and $\boldsymbol{B}_{N,W}\boldsymbol{x}$ has a fast implementation via the FFT, Chapter 3 provides an efficient method for solving the system $\boldsymbol{y} = \boldsymbol{B}_{N,W}\boldsymbol{x}$ which appears in linear prediction of bandlimited signals based on past samples and the Fourier extension.

CHAPTER 8

CONCLUSIONS AND POSSIBLE FUTURE WORK

This thesis considers parameterized subspace model in which the signals of interest are inherently low-dimensional and live in a union of subspaces, but the choice of subspace is controlled by a small number of continuous-valued parameters. Specifically, our focus has been constructing an appropriate basis (that matches the effective number of local degrees of freedom) to economically represent the signals of interest and developing rigorous, theoretically-backed techniques for computing projections onto and orthogonal to these subspaces. Our contributions in this thesis have included: new non-asymptotic results on the eigenvalue distribution of (periodic) discrete time-frequency localization operators; fast constructions for computing approximate projections onto the leading Slepian basis elements; an orthogonal approximate Slepian transform that has computational complexity comparable to the FFT; results on the spectrum of combined time- and multiband-limiting operations in the discrete-time domain and analysis for a dictionary formed by concatenating a collection of modulated DPSS's; analysis for the dimensionality of wall and target return subspaces in through-the-wall radar imaging and algorithms for mitigating wall clutter and identifying non-point targets; an asymptotic performance guarantee for the individual eigenvalue estimates for Toeplitz matrices by circulant matrices; and the eigenvalue distribution of time-frequency limiting operators on locally compact abelian groups. To conclude, we point out some ongoing research as well as many possible future directions for this research.

8.1 On the Asymptotic Equivalence of Circulant and Toeplitz Matrices

In Chapter 6, we took the first step towards investigating conditions under which the asymptotic equivalence of the matrices implies the individual asymptotic convergence of the eigenvalues. In particular, our results suggest that instead of directly computing the eigenvalues of a Toeplitz matrix, one can compute a fast spectrum approximation using the

FFT. This is long known, but we provide new guarantees for the asymptotic convergence of the individual eigenvalues.

However, the convergence rate $O(\frac{1}{N})$ (see Theorem 6.3) is only provided for band Toeplitz matrices. A number of issues remain open regarding the convergence rate of the eigenvalues between the general Toeplitz matrices and circulant matrices constructed in Chapter 6. For instance, we suspect a similar convergence rate ($O(\frac{1}{N})$ or $O(\frac{\log N}{N})$) holds under the assumptions of Theorem 6.2 (and even Theorem 6.4).

Another interesting question would be whether it is possible to extend our analysis to general (non-Hermitian) Toeplitz matrices, along the lines of the Avram-Parter theorem [6, 103]. In addition, it would also be of interest to extend our analysis to the asymptotic equivalence of block Toeplitz and block circulant matrices.

8.2 Time-Frequency Limiting Operators on Locally Compact Abelian Groups

In Chapter 7, we characterize the asymptotic behavior for the eigenvalues of the composite time- and frequency-limiting operators on locally compact abelian groups. As inspired by the applications listed in Section **??**, we raise several questions concerning Theorems 7.1 and 7.2. Following the two remarks after Theorem 7.1, two natural questions are raised as follows: (i) Can we improve the second term in (7.9)? Furthermore, what nonasymptotic result (like (7.14) for the PSWF eigenvalues and (7.15) for the DPSS eigenvalues) can we obtain for the number of eigenvalues of the Toeplitz operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ within the transition region $(\epsilon, 1-\epsilon)$? (ii) Can we extend (7.16) to the general time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$?

Other open theoretical questions concern the accuracy to which the subspace spanned by the first n eigenfunctions of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ can represent each individual time-limited character $\mathcal{T}_{\mathbb{A}}\chi_\xi$ with $\xi \in \mathbb{B}$. Theorem 7.2 ensures that accuracy is guaranteed in a MSE sense if one chooses $n \approx |\mathbb{A}||\mathbb{B}|$ such that the sum of the remaining eigenvalues of G is small. We suspect that a uniform guarantee for each $\mathcal{T}_{\mathbb{A}}\chi_\xi$ can also be obtained since the derivative of $\|\chi_\xi\|^2_{L_2(\mathbb{A})}$ is bounded, giving a finer result concerning the eigenvalue distribution for $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$. In Chapter 4, we have a non-asymptotic guarantee for the DPSS basis in representing each

complex exponential $\boldsymbol{e}_f$ with frequency f inside a band of interest.

8.3 The General Framework in the Discrete Case

Inspired by the results in Chapters 3, 4 and 5 respectively for sampled bandlimited signals and sampled multiband signals (see (2.5) in Section 2.4.2), we consider the general problem of finding appropriate subspace that best approximates the data vector $\boldsymbol{x}$ drawn from a collection of subspaces of $\mathbb{C}^N$ parameterized by a small number of continuous-valued parameters. The information content of such signals $\boldsymbol{x}$ is governed by the small number of continuous-valued parameters. The desired subspace is to materialize the intrinsic information content and facilitate the subsequent data processing tasks.

8.3.1 Formal Description

To begin, we let $\mathcal{S}_\theta$ denote an L-dimensional subspace of $\mathbb{C}^N$ parameterized by $\theta = (\theta_1, \theta_2, \ldots, \theta_D) \in \Theta$, where Θ is the D-dimensional parameter space that governs our signal. We let the matrix $\boldsymbol{\Psi}_\theta \in \mathbb{C}^{N\times L}$ denote any orthonormal basis[17] for $\mathcal{S}_\theta$. Note that $\boldsymbol{\Psi}_\theta$ can be viewed as a matrix-valued function of θ. We assume $\boldsymbol{\Psi}_\theta[n, l]$ is a continuously integrable function of θ for all $n \in [N], l \in [L]$. The signal of interest is generated as an *integral* of vectors belonging to different subspaces $\mathcal{S}_\theta$

$$\boldsymbol{x} = \int_{\mathbb{D}\subset\Theta} \boldsymbol{v}_\theta d\theta \quad \text{where} \quad \boldsymbol{v}_\theta = \boldsymbol{\Psi}_\theta \boldsymbol{\alpha}_\theta \in \mathcal{S}_\theta, \tag{8.1}$$

where $\mathbb{D} \subset \Theta$ represents a possibly infinite set of parameter vectors and $\{\boldsymbol{\alpha}_\theta\}_{\theta\in\Theta}$ are the length-L vector valued functions of coefficients corresponding to the bases $\{\mathcal{S}_\theta\}_{\theta\in\Theta}$.

In many applications, $\mathbb{D}$ will have small measure, and we are interesting in several questions:

- what is an appropriate low-dimensional basis that best approximates all possible $\boldsymbol{x}$ appearing in (8.1)?
- how do the dimensionality and basis functions change as a function of $\mathbb{D}$?

[17] This condition maybe reduced to just basis.

8.3.2 Potential Approach

In order to answer the questions posted above, similar to the approach in Section 7.1.2, we may define an operator $\mathcal{A} : \mathbb{C}^N \to L_2(\mathbb{D}, \mathbb{C}^L)$ as

$$[\mathcal{A}\boldsymbol{u}](\theta) = \boldsymbol{\Psi}_\theta^H \boldsymbol{u}, \quad \theta \in \mathbb{D},$$

where $[\mathcal{A}\boldsymbol{u}](\theta)$ is a vector valued function. Define the vector valued function $\boldsymbol{a}_\theta : \mathbb{R}^D \to \mathbb{C}^L$ via

$$\boldsymbol{a}_\theta = \left[a_0(\theta)\ a_1(\theta)\ \cdots\ a_{L-1}(\theta)\right]^T,$$

where $a_l(\theta) \in C$ is a function of θ for all $l \in [L]$. Here the normed linear space $L_2(\mathbb{D}, \mathbb{C}^L)$ consists of all vector valued functions $\boldsymbol{a}_\theta$ on the interval $\theta \in \mathbb{D}$ with

$$\|\boldsymbol{a}\|^2_{L_2(\mathbb{D},\mathbb{C}^L)} = \int_{\mathbb{D}} \|\boldsymbol{a}_\theta\|_2^2 d\theta < \infty.$$

Letting $\boldsymbol{b}_\theta \in L_2(\mathbb{D}, \mathbb{C}^L)$, we can also define the inner product in $L_2(\mathbb{D}, \mathbb{C}^L)$ via

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{L_2(\mathbb{D},\mathbb{C}^L)} = \int_{\mathbb{D}} \langle \boldsymbol{a}_\theta, \boldsymbol{b}_\theta \rangle\, d\theta = \int_{\mathbb{D}} \boldsymbol{b}_\theta^H \boldsymbol{a}_\theta d\theta.$$

The adjoint $\mathcal{A}^* : L_2(\mathbb{D}, \mathbb{C}^L) \to \mathbb{C}^N$ can be expressed as

$$\mathcal{A}^* \boldsymbol{a} = \int_{\mathbb{D}} \boldsymbol{\Psi}_\theta \boldsymbol{a}_\theta\ d\theta.$$

Thus the signal model appearing in (8.1) equals $\boldsymbol{x} = \mathcal{A}^* \boldsymbol{a}$. One can check that

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{a} \rangle_{L_2(\mathbb{D},\mathbb{C}^L)} = \int_\tau \boldsymbol{a}_\theta^H \langle \boldsymbol{u}, \boldsymbol{\Psi}_\theta \rangle d\theta = \int_\tau \boldsymbol{a}_\theta^H \boldsymbol{\Psi}_\theta^H \boldsymbol{u} d\theta = \langle \boldsymbol{u}, \mathcal{A}^* \boldsymbol{a} \rangle,$$

where the notation $\langle \cdot, \cdot \rangle$, without subscript, denotes the conventional inner product on $\mathbb{C}^N$. The self-adjoint operator $\mathcal{A}^*\mathcal{A} : \mathbb{C}^N \to \mathbb{C}^N$ is given by

$$\mathcal{A}^*\mathcal{A}\boldsymbol{u} = \int_{\mathbb{D}} \boldsymbol{\Psi}_\theta \boldsymbol{\Psi}_\theta^H \boldsymbol{u} d\theta = \int_{\mathbb{D}} \boldsymbol{\Psi}_\theta \boldsymbol{\Psi}_\theta^H d\theta \boldsymbol{u} = \boldsymbol{B}\boldsymbol{u} \tag{8.2}$$

with

$$\boldsymbol{B} = \int_{\mathbb{D}} \boldsymbol{\Psi}_\theta \boldsymbol{\Psi}_\theta^H d\theta, \tag{8.3}$$

which is equivalent to $\mathcal{A}^*\mathcal{A}$. We can show that $\boldsymbol{B}$ is the covariance matrix of a random variable generated by picking θ uniformly from $\mathbb{D}$, and thus investigating the operator $\mathcal{A}^*\mathcal{A}$ is equivalent to the conventional PCA approach. Similarly, we can define the operator $\mathcal{A}\mathcal{A}^*$ which is a Hilbert-Schmidt integral operator and has at most N non-zero eigenvalues. One can check that $\mathcal{A}^*\mathcal{A}$ and $\mathcal{A}\mathcal{A}^*$ have the same non-zero eigenvalues.

DPSS and DPSWF Revisited

Consider the example where $\mathbb{D} = [-W, W] \subset \Theta = [-\frac{1}{2}, \frac{1}{2}]$ and $L = 1$, $\boldsymbol{\Psi}_\theta[n] = e^{j2\pi n}, n \in [N]$. The corresponding signal class consists of the sampled bandlimited signals as

$$\boldsymbol{x}[n] = \int_{\mathbb{D}} \alpha_\theta e^{j2\pi\theta n} d\theta.$$

The operator $\mathcal{A} : \mathbb{C}^N \to L_2(\mathbb{D})$ becomes

$$[\mathcal{A}\boldsymbol{u}](\theta) = \sum_{n=0}^{N-1} e^{-j2\pi\theta n} \boldsymbol{u}[n], \ \theta \in \mathbb{D}.$$

In words, the operator $\mathcal{A}$ is equivalent to taking a vector in $\mathbb{C}^N$, computing its DTFT and truncating the DTFT to $\mathbb{D}$. The adjoint operator $\mathcal{A}^* : L_2(\mathbb{D}) \to \boldsymbol{C}^N$ acts on a function supported on $\mathbb{D}$, computing its inverse DTFT (IDTFT), and then time-limiting the IDTFT to $[N]$ as

$$[\mathcal{A}^* a_\theta][n] = \int_{\mathbb{D}} a_\theta e^{j2\pi\theta n} d\theta, \ n \in [N].$$

We then obtain the self-adjoint operator $\mathcal{A}^*\mathcal{A} : \mathbb{C}^N \to \mathbb{C}^N$

$$\mathcal{A}^*\mathcal{A}\boldsymbol{u} = \boldsymbol{B}_{N,W}\boldsymbol{u}$$

where $\boldsymbol{B}_{N,W}$ is the prolate matrix defined in (2.16) (and also (2.11)). The operator $\mathcal{A}^*\mathcal{A}$ is equivalent to the composite of time- and band-limiting operators in Section 2.5. Thus the eigenfunctions of $\mathcal{A}^*\mathcal{A}$ are the DPSS's.

In a nutshell, the number of large eigenvalues of $\mathcal{A}^*\mathcal{A}$ tells the effective dimensionality of such signals $\boldsymbol{x}$ modelled in (8.1). The eigenfunctions (or eigenvectors) of $\mathcal{A}^*\mathcal{A}$ corresponding to the significant eigenvalues give us an orthobasis for the class signal $\boldsymbol{x}$ appearing in (8.1). Thus, to answer the questions posted in Section 8.3.1, we can turn to investigate the eigenvalues of either $\mathcal{A}^*\mathcal{A}$ or $\mathcal{A}\mathcal{A}^*$. In some cases a closed form expression for matrix $\boldsymbol{B}$ may be available; in other cases this matrix may be computable (e.g., via integration of $\int_{\mathbb{D}} \boldsymbol{\Psi}_\theta \boldsymbol{\Psi}_\theta^H d\theta$). Numerical integration for $\int_{\mathbb{D}} \boldsymbol{\Psi}_\theta \boldsymbol{\Psi}_\theta^H d\theta$ can be obtained by uniformly sampling θ over $\mathbb{D}$ (for smoothness class of integrands) and by randomly sampling θ over $\mathbb{D}$ (for complicated integrands). Once the matrix $\boldsymbol{B}$ is a Toeplitz matrix, we can utilize the fast estimates developed in Chapter 6 for estimating the eigenvalues of $\boldsymbol{B}$. However, a number of theoretical questions still remain open concerning the eigenvalue distribution for the operators $\mathcal{A}^*\mathcal{A}$ or $\mathcal{A}\mathcal{A}^*$, like Corollary 3.1, Theorem 5.3 and Theorem 7.1 for the eigenvalue distribution of the time-frequency limiting operators.

REFERENCES CITED

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.

[2] F. Ahmad, Q. Jiang, and M. Amin. Wall clutter mitigation using discrete prolate spheroidal sequences for sparse reconstruction of indoor stationary scenes. *IEEE Trans. Geosci. Remote Sens.*, 53(3):1549–1557, 2015.

[3] N. Ailon and B. Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on computing*, 39(1):302–322, 2009.

[4] G. Ammar and W. Gragg. Superfast solution of real positive definite Toeplitz systems. *SIAM J. Matrix Anal. Appl.*, 9(1):61–76, 1988.

[5] T. M. Apostol. *Mathematical Analysis (2nd ed.)*. Addison Wesley Publishing Company, 1974.

[6] F. Avram. On bilinear forms in gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields*, 79(1):37–45, 1988.

[7] R. G. Baraniuk and P. Steeghs. Compressive radar imaging. In *Proc. 2007 IEEE Radar Conference*, pages 128–133, April 2007.

[8] M. V. Barel, G. Heinig, and P. Karvanja. A superfast method for solving Toeplitz linear least squares problems. *Linear Algebra Appl.*, 366:441–457, 2003.

[9] R. Bitmead and B. Anderson. Asymptotically fast solutions of Toeplitz and related systems of equations. *Linear Algebra Appl.*, 34:103–116, 1980.

[10] A. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of computation*, 27(123):579–594, 1973.

[11] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, 2008.

[12] J. Bogoya, A. Böttcher, S. Grudsky, and E. Maximenko. Maximum norm versions of the Szegő and Avram-Parter theorems for Toeplitz matrices. *J. Approx. Theory*, 196(0):79 – 100, 2015.

[13] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[14] P. A. Businger and G. H. Golub. Algorithm 358: Singular value decomposition of a complex matrix [F1, 4, 5]. *Comm. ACM*, 12(10):564–565, 1969.

[15] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[16] E. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

[17] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. Pure Appl. Math.*, 67(6):906–956, 2014.

[18] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[19] R. H. Chan. Circulant preconditioners for hermitian Toeplitz systems. *SIAM J. Matrix Anal. Appl.*, 10(4):542–550, 1989.

[20] R. H. Chan, X.-Q. Jin, and M.-C. Yeung. The spectra of super-optimal circulant preconditioned Toeplitz systems. *SIAM J. Numer. Anal.*, 28(3):871–879, 1991.

[21] R. H. Chan and G. Strang. Toeplitz equations by conjugate gradients with circulant preconditioner. *SIAM J. Sci. Statis. Comput.*, 10(1):104–119, 1989.

[22] T. F. Chan. An optimal circulant preconditioner for Toeplitz systems. *SIAM J. Sci. Statis. Comput.*, 9(4):766–771, 1988.

[23] C.-Y. Chen and P. P. Vaidyanathan. Mimo radar space–time adaptive processing using prolate spheroidal wave functions. *IEEE Trans. Signal Process.*, 56(2):623–635, 2008.

[24] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

[25] G. S. Chirikjian and A. B. Kyatkin. *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition*. Courier Dover Publications, 2016.

[26] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

[27] D. D. Cox. Spectral analysis for physical applications: Multitaper and conventional univariate techniques, 1996.

[28] F. Dahlen and F. J. Simons. Spectral estimation on a sphere in geophysics and cosmology. *Geophysical Journal International*, 174(3):774–807, 2008.

[29] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk. Signal processing with compressive measurements. *IEEE J. Select. Top. Signal Processing*, 4(2):445–460, 2010.

[30] M. Davenport, S. Schnelle, J. Slavinsky, R. Baraniuk, M. Wakin, and P. Boufounos. A wideband compressive radio receiver. In *Military Communications Conference (MILCOM)*, pages 1193–1198, Oct 2010.

[31] M. Davenport, S. Schnelle, J. P. Slavinsky, R. Baraniuk, M. Wakin, and P. Boufounos. A wideband compressive radio receiver. In *Proc. Military Comm. Conf. (MILCOM)*, San Jose, California, Oct. 2010.

[32] M. Davenport and M. Wakin. Compressive sensing of analog signals using discrete prolate spheroidal sequences. *Appl. Comput. Harmon. Anal.*, 33(3):438–472, 2012.

[33] M. A. Davenport and M. B. Wakin. Reconstruction and cancellation of sampled multiband signals using discrete prolate spheroidal sequences. In *Proc of Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS11)*, page 61, 2011.

[34] G. Davis. *Adaptive nonlinear approximations*. PhD thesis, Courant Institute of Mathematical Sciences New York, 1994.

[35] P. Deift, A. Its, and I. Krasovsky. Eigenvalues of Toeplitz matrices in the bulk of the spectrum. *Bull. Inst. Math., Acad. Sin. (N.S.)*, pages 437–461, 2012.

[36] A. Dembo. Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices. *IEEE Trans. Inf. Theory*, 34(2):352–355, 1988.

[37] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[38] G. T. Di Francia. Degrees of freedom of an image. *J. Opt. Soc. Am.*, 59(7):799–804, 1969.

[39] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*, 41(3):613–627, 1995.

[40] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[41] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. *Proc. Natt. Acad. Sci.*, 100(5):2197–2202, 2003.

[42] W. Du, G. Liao, and Z. Yang. Robust space time processing based on bi-iterative scheme of secondary data selection and pswf method. *Digital Signal Processing*, 52:64–71, 2016.

[43] B. Duggal. Subspace gaps and range-kernel orthogonality of an elementary operator. *Linear algebra and its applications*, 383:93–106, 2004.

[44] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[45] A. Edelman, P. McCorquodale, and S. Toledo. The future fast Fourier transform? *SIAM J. Sci. Comput.*, 20(3):1094–1114, 1998.

[46] A. Eftekhari, J. Romberg, and M. B. Wakin. Matched filtering from limited frequency samples. *IEEE Trans. Inf. Theory*, 59(6):3475–3496, 2013.

[47] A. Eftekhari and M. B. Wakin. Greed is super: A fast algorithm for super-resolution. *arXiv preprint arXiv:1511.03385*, 2015.

[48] K. Engan, S. O. Aase, and J. Hakon Husoy. Method of optimal directions for frame design. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, volume 5, pages 2443–2446, 1999.

[49] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Imaging Sci.*, 5(1):179–202, 2012.

[50] M. Franceschetti. On landau's eigenvalue theorem and its applications. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 381–385. IEEE, 2014.

[51] M. Franceschetti. On Landau's eigenvalue theorem and information cut-sets. *IEEE Trans. Inform. Theory*, 61(9):5042–5051, 2015.

[52] H. Gazzah, P. A. Regalia, and J.-P. Delmas. Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification. *IEEE Trans. Inf. Theory*, 47(3):1243–1251, 2001.

[53] L. Gosse. Compressed sensing with preconditioning for sparse recovery with subsampled matrices of slepian prolate functions. *ANNALI DELL'UNIVERSITA'DI FERRARA*, 59(1):81–116, 2013.

[54] R. Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Trans. Inf. Theory*, 18(6):725–730, 1972.

[55] R. M. Gray. Toeplitz and circulant matrices: A review. *Commun. Inf. Theory*, 2(3):155–239, 2005.

[56] U. Grenander and G. Szegő. *Toeplitz Forms and Their Applications*, volume 321. Univ of California Press, 1958.

[57] F. A. Grünbaum. Eigenvectors of a toeplitz matrix: discrete version of the prolate spheroidal wave functions. *SIAM J. Algeb. Discrete Methods*, 2(2):136–141, 1981.

[58] F. A. Grünbaum. Toeplitz matrices commuting with tridiagonal matrices. *Linear Algebra Appl.*, 40(0):25 – 36, 1981.

[59] J. Gutiérrez-Gutiérrez and P. M. Crespo. Asymptotically equivalent sequences of matrices and hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems. *IEEE Trans. Inf. Theory*, 54(12):5671–5680, 2008.

[60] N. Halko, P. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53(2):217–288, 2011.

[61] S. Haykin. Cognitive radio: Brain-empowered wireless communications. *IEEE J. Select. Areas Commun.*, 23(2):201–220, 2005.

[62] G. Heinig. Inversion of generalize Cauchy matrices and other classes of structured matrices. In *Linear Algebra for Signal Processing*. Springer, 1995.

[63] J. A. Hogan and J. D. Lakey. *Duration and Bandwidth Limiting: Prolate Functions, Sampling, and Applications*. Springer Science & Business Media, 2011.

[64] J. A. Hogan and J. D. Lakey. Frame properties of shifts of prolate spheroidal wave functions. *Appl. Comput. Harmonic Anal.*, 39(1):21–32, 2015.

[65] J. A. Hogan and J. D. Lakey. Wavelet frames generated by bandpass prolate functions. In *Int. Conf. Sampling Theory Applic. (SampTA)*, pages 120–123. IEEE, 2015.

[66] J. M. Hokanson. Projected nonlinear least squares for exponential fitting. *arXiv preprint arXiv:1508.05890*, 2015.

[67] R. A. Horn and C. R. Johnson, editors. *Matrix Analysis*. Cambridge University Press, New York, NY, USA, 1986.

[68] D. Huybrechs. On the Fourier extension of nonperiodic functions. *SIAM J. Numer. Anal.*, 47(6):4326–4355, 2010.

[69] A. Israel. The eigenvalue distribution of time-frequency localization operators. *arXiv:1502.04404*, 2015.

[70] S. Izu and J. D. Lakey. Time-frequency localization and sampling of multiband signals. *Acta Appl. Math.*, 107(1-3):399–435, 2009.

[71] A. Jain and S. Ranganath. Extrapolation algorithms for discrete signals with application in spectral estimation. *IEEE Trans. Acous. Speech Signal Process.*, 29(4):830–845, 1981.

[72] T. Kailath, S. Kung, and M. Morf. Displacement ranks of matrices and linear equations. *J. Math. Anal. Appl.*, 68(2):395–407, 1979.

[73] T. Kailath, A. Vieira, and M. Morf. Inverses of Toeplitz operators, innovations, and orthogonal polynomials. *SIAM Rev.*, 20(1):106–119, 1978.

[74] S. Karnik, Z. Zhu, M. Wakin, J. Romberg, and M. Davenport. Fast computations for approximation and compression in Slepian spaces. In *Proc. IEEE. Global Conf. on Signal and Inform. Processing (GlobalSIP)*, Washington, DC, Dec. 2016.

[75] S. Karnik, Z. Zhu, M. B. Wakin, J. K. Romberg, and M. A. Davenport. The fast Slepian transform. to appear in *Appl. Comp. Harm. Anal., arXiv preprint arXiv:1611.04950*.

[76] A. Kolmogorov. Über die beste anna′aherung von funktionen einer gegebenen funktionenklasse. *Annals of Mathematics*, 37(1):107–110, 1936.

[77] T. W. Körner. *Fourier analysis*. Cambridge university press, 1989.

[78] E. Lagunas, M. G. Amin, F. Ahmad, and M. Najar. Joint wall mitigation and compressive sensing for indoor image reconstruction. *IEEE Trans. Geosci. Remote Sens.*, 51(2):891–906, 2013.

[79] H. Landau. On szegőŠs eingenvalue distribution theorem and non-hermitian kernels. *Journal d'Analyse Mathématique*, 28(1):335–357, 1975.

[80] H. Landau. On the density of phase-space expansions. *IEEE Trans. Inform. Theory*, 39(4):1152–1156, 1993.

[81] H. Landau and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. II. *Bell Systems Tech. J.*, 40(1):65–84, 1961.

[82] H. Landau and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. III. *Bell Systems Tech. J.*, 41(4):1295–1336, 1962.

[83] H. Landau and H. Widom. Eigenvalue distribution of time and frequency limiting. *J. Math. Anal. Appl.*, 77(2):469–481, 1980.

[84] T. Laudadio, N. Mastronardi, and M. Van Barel. Computing a lower bound of the smallest eigenvalue of a symmetric positive-definite Toeplitz matrix. *IEEE Trans. Inf. Theory*, 54(10):4726–4731, 2008.

[85] D. Lawden. *Elliptic Functions and Applications*. Springer-Verlag, 1989.

[86] R. Lenz. Group-theoretical model of feature extraction. *J. Opt. Soc. Am.*, 6(6):827–834, 1989.

[87] G. Li, Z. Zhu, H. Bai, and A. Yu. A new framework for designing incoherent sparsifying dictionaries. In *IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, pages 4416–4420, 2017.

[88] J.-R. Li and J. White. Low rank solution of Lyapunov equations. *SIAM J. Matrix Anal. Appl.*, 24(1):260–280, 2002.

[89] T. J. Lim and M. Franceschetti. Information without rolling dice. *IEEE Trans. Inform. Theory*, 2017.

[90] A. Lu and E. Wachspress. Solution of Lyapunov equations by alternating direction implicit iteration. *Comput. Math. Appl.*, 21(9):43–58, 1991.

[91] J. Makhoul. Linear prediction: A tutorial review. *Proc. IEEE*, 63(4):561–580, 1975.

[92] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 3rd edition, 2008.

[93] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.

[94] P. Martinsson, V. Rokhlin, and M. Tygert. A fast algorithm for the inversion of general Toeplitz matrices. *Comput. Math. Appl.*, 50(5-6):741–752, 2005.

[95] R. Matthysen and D. Huybrechs. Fast algorithms for the computation of Fourier extensions of arbitrary length. *SIAM J. Sci. Comput.*, 38(2):A899–A922, 2016.

[96] L. Mirsky. A trace inequality of john von neumann. *Monatshefte für Mathematik*, 79(4):303–306, 1975.

[97] J. Mitola III and G. Q. Maguire Jr. Cognitive radio: Making software radios more personal. *IEEE Personal Commun.*, 6(4):13–18, 1999.

[98] M. Morf. *Fast algortihms for multivariable systems*. PhD thesis, Stanford University, 1974.

[99] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmonic Anal.*, 26(3):301–321, 2009.

[100] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, USA, 1st edition, 2010.

[101] A. Osipov. Certain upper bounds on the eigenvalues associated with prolate spheroidal wave functions. *Appl. Comput. Harmon. Anal.*, 35(2):309–340, 2013.

[102] V. Pan. *Structured matrices and polynomials: Unified superfast algorithms*. Birkhauser/Springer, 2001.

[103] S. V. Parter. On the distribution of the singular values of Toeplitz matrices. *Linear Algebra Appl.*, 80:115–130, 1986.

[104] J. Pearl. On coding and filtering stationary signals by discrete Fourier transforms. *IEEE Trans. Inf. Theory*, 19(2):229–232, 1973.

[105] L. Reichel and L. N. Trefethen. Eigenvalues and pseudo-eigenvalues of Toeplitz matrices. *Linear Algebra Appl.*, 162:153–185, 1992.

[106] W. Rudin. *Fourier Analysis on Groups*. John Wiley & Sons, 2011.

[107] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[108] D. Sakrison. An extension of the theorem of Kac, Murdock and Szegő to N dimensions. *IEEE Trans. Inf. Theory*, 15(5):608–610, 1969.

[109] I. Schur. Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *J. Reine Angew. Math.*, 140:1–28, 1911.

[110] E. Sejdić, A. Can, L. Chaparro, C. Steele, and T. Chau. Compressive sampling of swallowing accelerometry signals using time-frequency dictionaries based on modulated Discrete Prolate Spheroidal Sequences. *EURASIP J. Adv. Signal Processing*, 2012(1):1–14, 2012.

[111] E. Sejdić, M. Luccini, S. Primak, K. Baddour, and T. Willink. Channel estimation using DPSS based frames. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Las Vegas, NV, Mar. 2008.

[112] S. Senay, L. F. Chaparro, and L. Durak. Reconstruction of nonuniformly sampled time-limited signals using prolate spheroidal wave functions. *Signal Process.*, 89(12):2585–2595, 2009.

[113] S. Senay, L. F. Chaparro, M. Sun, and R. J. Sclabassi. Compressive sensing and random filtering of eeg signals using slepian basis. In *16th European Signal Processing Conference*, pages 1–5, 2008.

[114] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[115] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. IV. *Bell Systems Tech. J.*, 43(6):3009–3058, 1964.

[116] D. Slepian. Some asymptotic expansions for prolate spheroidal wave functions. *J. of Math. Physics*, 44(1):99–140, 1965.

[117] D. Slepian. On Bandwidth. *Proc. IEEE*, 64(3):292–300, 1976.

[118] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V – The discrete case. *Bell Systems Tech. J.*, 57(5):1371–1430, 1978.

[119] D. Slepian. Some comments on Fourier analysis, uncertainty, and modeling. *SIAM Rev.*, 25(3):379–393, 1983.

[120] D. Slepian and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. I. *Bell Systems Tech. J.*, 40(1):43–64, 1961.

[121] H. Stark and J. W. Woods. *Probability, Random Processes, and Estimation Theory for Engineers*. Prentice-Hall, 1986.

[122] G. Strang. A proposal for Toeplitz matrix calculations. *Stud. Appl. Math.*, 74(2):171–176, 1986.

[123] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere. *arXiv preprint arXiv:1504.06785*, 2015.

[124] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Inf. Theory*, 59(11):7465–7490, 2013.

[125] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70(9):1055–1096, 1982.

[126] W. F. Trench. An elementary view of Weyl's theory of equal distribution. *Amer. Math. Monthly*, 119(10):852–861, 2012.

[127] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.

[128] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond Nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Trans. Inf. Theory*, 56(1):520–544, 2010.

[129] C. Turnes. *Efficient Solution to Toeplitz-Structured Linear Systems for Signal Processing*. PhD thesis, Georgia Institute of Technology, 2014.

[130] E. E. Tyrtyshnikov. A unifying approach to some old and new theorems on distribution and clustering. *Linear Algebra Appl.*, 232:1–43, 1996.

[131] A. W. Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge university press, 2000.

[132] J. Varah. The prolate matrix. *Linear Algebra Appl.*, 187:269–278, 1993.

[133] C. Vonesch, F. Stauber, and M. Unser. Steerable pca for rotation-invariant image recognition. *SIAM Journal on Imaging Sciences*, 8(3):1857–1873, 2015.

[134] E. Wachspress. *The ADI Model Problem*. Springer-Verlag, 2013.

[135] H. Xiao, V. Rokhlin, and N. Yarvin. Prolate spheroidal wavefunctions, quadrature and interpolation. *Inverse problems*, 17(4):805, 2001.

[136] J. Xu. Degrees of freedom of oam-based line-of-sight radio systems. *IEEE Transactions on Antennas and Propagation*, 65(4):1996–2008, 2017.

[137] W. Y. Xu and C. Chamzas. On the periodic discrete prolate spheroidal sequences. *SIAM J. Appl. Math.*, 44(6):1210–1217, 1984.

[138] X. Yang, Y. Liu, and T. Long. Robust non-homogeneity detection algorithm based on prolate spheroidal wave functions for space-time adaptive processing. *IET Radar, Sonar & Navigation*, 7(1):47–54, 2013.

[139] N. L. Zamarashkin and E. E. Tyrtyshnikov. Distribution of eigenvalues and singular values of Toeplitz matrices under weakened conditions on the generating function. *Sbornik: Mathematics*, 188(8):1191–1201, 1997.

[140] T. Zemen and C. Mecklenbräuker. Time-variant channel estimation using Discrete Prolate Spheroidal Sequences. *IEEE Trans. Signal Processing*, 53(9):3597–3607, 2005.

[141] T. Zemen, C. Mecklenbräuker, F. Kaltenberger, and B. Fleury. Minimum-energy band-limited predictor with dynamic subspace selection for time-variant flat-fading channels. *IEEE Trans. Signal Processing*, 55(9):4534–4548, 2007.

[142] Y. Zeng and Y.-C. Liang. Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Trans Commun.*, 57(6):1784–1793, 2009.

[143] Z. Zhu, S. Karnik, M. A. Davenport, J. K. Romberg, and M. B. Wakin. The eigenvalue distribution of discrete periodic time-frequency limiting operators. *to appear in IEEE Signal Process. Letters*, 2017.

[144] Z. Zhu, S. Karnik, M. B. Wakin, M. A. Davenport, and J. K. Romberg. Fast orthogonal approximations of sampled sinusoids and bandlimited signals. In *IEEE Conf. Acous., Speech, Signal Process. (ICASSP)*, pages 4511–4515, 2017.

[145] Z. Zhu and M. Wakin. Wall clutter mitigation and target detection using Discrete Prolate Spheroidal Sequences. In *Proc. Int. Work. on Compressed Sensing Theory Appl. Radar, Sonar and Remote Sens. (CoSeRa)*, Pisa, Italy, Jun. 2015.

[146] Z. Zhu and M. B. Wakin. Detection of stationary targets using Discrete Prolate Spheroidal Sequences. In *31st International Review of Progress in Applied Computational Electromagnetics (ACES)*, pages 1–2, March 2015.

[147] Z. Zhu and M. B. Wakin. On the dimensionality of wall and target return subspaces in through-the-wall radar imaging. In *4th Int. Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, September 2016.

[148] Z. Zhu and M. B. Wakin. Approximating sampled sinusoids and multiband signals using multiband modulated DPSS dictionaries. *J. Fourier Anal. Appl.*, 23(6):1263–1310, Dec 2017.

[149] Z. Zhu and M. B. Wakin. On the asymptotic equivalence of circulant and Toeplitz matrices. *IEEE Trans. Inform. Theory*, 63(5):2975–2992, 2017.

[150] P. Zizler, R. A. Zuidwijk, K. F. Taylor, and S. Arimoto. A finer aspect of eigenvalue distribution of selfadjoint band Toeplitz matrices. *SIAM J. Matrix Anal. Appl.*, 24(1):59–67, 2002.

APPENDIX A

PROOFS FOR CHAPTER 3

This appendix mainly includes the proofs for Chapter 3.

A.1 Proof of Theorem 3.1

Our goal is to show that $\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ is well-approximated as a factored low rank matrix. To do this, we will express $\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ in terms of other matrices, whose entries also have a closed form. We will then derive a factored low rank approximation for each of these other matrices. Finally, we will combine these low rank approximations to get a factored low rank approximation for $\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$.

Towards this end, we define $\boldsymbol{D}_A$ to be the $N \times N$ diagonal matrix with diagonal entries $\boldsymbol{D}_A[n,n] = e^{j2\pi W' n}$ for $n = 0, \ldots, N-1$ and define $\boldsymbol{A}_0$ to be the $N \times N$ matrix with entries

$$\boldsymbol{A}_0[m,n] = \begin{cases} \frac{1}{\pi(m-n)} - \frac{1}{N\sin\left(\pi\frac{m-n}{N}\right)} & \text{if } m \neq n, \\ 0 & \text{if } m = n. \end{cases}$$

We also define $\boldsymbol{D}_B$ to be the $N \times N$ diagonal matrix with diagonal entries $\boldsymbol{D}_B[n,n] = e^{j(W+W')n}$ for $n = 0, \ldots, N-1$ and define $\boldsymbol{B}_0$ to be the $N \times N$ matrix with entries

$$\boldsymbol{B}_0[m,n] = \frac{2\sin(\pi(W+W')(m-n))}{\pi(m-n)}.$$

Note that with these definitions, we can write the (m,n)-th entry of $\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ as

$$
\begin{aligned}
&[\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*][m,n] \\
&= \frac{\sin(2\pi W(m-n))}{\pi(m-n)} - \frac{\sin(2\pi W'(m-n))}{N\sin(\pi\frac{m-n}{N})} \\
&= \frac{\sin(2\pi W(m-n))}{\pi(m-n)} - \frac{\sin(2\pi W'(m-n))}{\pi(m-n)} + \frac{\sin(2\pi W'(m-n))}{\pi(m-n)} - \frac{\sin(2\pi W'(m-n))}{N\sin(\pi\frac{m-n}{N})} \\
&= \frac{2\sin(\pi(W-W')(m-n))\cos(\pi(W+W')(m-n))}{\pi(m-n)} \\
&\quad + \frac{\sin(2\pi W'(m-n))}{\pi(m-n)} - \frac{\sin(2\pi W'(m-n))}{N\sin(\pi\frac{m-n}{N})} \\
&= \boldsymbol{B}_0[m,n]\cos(\pi(W+W')(m-n)) + \boldsymbol{A}_0[m,n]\sin(2\pi W'(m-n)) \\
&= \left[\frac{1}{2}\boldsymbol{D}_B\boldsymbol{B}_0\boldsymbol{D}_B^* + \frac{1}{2}\boldsymbol{D}_B^*\boldsymbol{B}_0\boldsymbol{D}_B + \frac{1}{2j}\boldsymbol{D}_A\boldsymbol{A}_0\boldsymbol{D}_A^* - \frac{1}{2j}\boldsymbol{D}_A^*\boldsymbol{A}_0\boldsymbol{D}_A\right][m,n].
\end{aligned}
$$

Hence,

$$
\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^* = \frac{1}{2j}\boldsymbol{D}_A\boldsymbol{A}_0\boldsymbol{D}_A^* - \frac{1}{2j}\boldsymbol{D}_A^*\boldsymbol{A}_0\boldsymbol{D}_A + \frac{1}{2}\boldsymbol{D}_B\boldsymbol{B}_0\boldsymbol{D}_B^* + \frac{1}{2}\boldsymbol{D}_B^*\boldsymbol{B}_0\boldsymbol{D}_B. \quad \text{(A.1)}
$$

Thus, we can find a low-rank approximation for $\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ by finding low-rank approximations for $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$. In order to do so, it is useful to consider the $N \times N$ matrix $\boldsymbol{A}_1$ defined by

$$
\boldsymbol{A}_1[m,n] = \begin{cases} \boldsymbol{A}_0[m,n] - \frac{1}{\pi(m-n+N)} - \frac{1}{\pi(m-n-N)} & \text{if } m \neq n, \\ 0 & \text{if } m = n. \end{cases}
$$

Next, we let $\boldsymbol{H}$ denote the *Hilbert matrix*, i.e., the $N \times N$ matrix with entries

$$
\boldsymbol{H}[m,n] = \frac{1}{m+n+1},
$$

and let $\boldsymbol{J}$ be the $N \times N$ matrix with 1's along the antidiagonal and zeros elsewhere. (Note that for an arbitrary $N \times N$ matrix $\boldsymbol{X}$, $\boldsymbol{J}\boldsymbol{X}$ is simply $\boldsymbol{X}$ flipped vertically and $\boldsymbol{X}\boldsymbol{J}$ is $\boldsymbol{X}$ flipped horizontally.) Using these definitions, we can write $\boldsymbol{A}_0$ as

$$
\boldsymbol{A}_0 = \frac{1}{\pi}(\boldsymbol{H}\boldsymbol{J} - \boldsymbol{J}\boldsymbol{H}) + \boldsymbol{A}_1. \quad \text{(A.2)}
$$

By combining (A.1) and (A.2), we get

$$\begin{aligned}\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^* =& \tfrac{1}{2j}\boldsymbol{D}_A\left[\tfrac{1}{\pi}(\boldsymbol{H}\boldsymbol{J}-\boldsymbol{J}\boldsymbol{H})+\boldsymbol{A}_1\right]\boldsymbol{D}_A^* - \tfrac{1}{2j}\boldsymbol{D}_A^*\left[\tfrac{1}{\pi}(\boldsymbol{H}\boldsymbol{J}-\boldsymbol{J}\boldsymbol{H})+\boldsymbol{A}_1\right]\boldsymbol{D}_A \\ &+ \tfrac{1}{2}\boldsymbol{D}_B\boldsymbol{B}_0\boldsymbol{D}_B^* + \tfrac{1}{2}\boldsymbol{D}_B^*\boldsymbol{B}_0\boldsymbol{D}_B.\end{aligned} \tag{A.3}$$

We next proceed by controlling the rank of $\boldsymbol{H}$, $\boldsymbol{A}_1$, and $\boldsymbol{B}_0$ separately.

Controlling the rank of $\boldsymbol{H}$

Our goal is to construct a low-rank matrix $\widetilde{\boldsymbol{H}} = \boldsymbol{Z}\boldsymbol{Z}^*$ such that $\|\boldsymbol{H} - \widetilde{\boldsymbol{H}}\| \leq \delta_H$ for some desired $\delta_H > 0$. We will do this via Lemma A.1, which we prove in Section A.6.

Lemma A.1. *Let $\boldsymbol{A}$ be an $N \times N$ symmetric positive definite matrix with condition number κ, let $\boldsymbol{B}$ be an arbitrary $N \times M$ matrix with $M \leq N$, and let $\boldsymbol{X}$ be the $N \times N$ positive definite solution to the Lyapunov equation*

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{A}^* = \boldsymbol{B}\boldsymbol{B}^*.$$

Then for any $\delta \in (0, 1]$, there exists an $N \times rM$ matrix $\boldsymbol{Z}$ with

$$r = \left\lceil \frac{1}{\pi^2}\log(4\kappa)\log\left(\frac{4}{\delta}\right)\right\rceil, \tag{A.4}$$

such that

$$\|\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{Z}^*\| \;\leq\; \delta\|\boldsymbol{X}\|. \tag{A.5}$$

Now, we let $\boldsymbol{A}$ be the $N \times N$ diagonal matrix defined by $\boldsymbol{A}[n,n] = n + \frac{1}{2}$ for $n = 0, \ldots, N-1$, and let $\boldsymbol{B} \in \mathbb{R}^N$ be a vector of all ones. It is easy to verify that the positive definite solution $\boldsymbol{X}$ to $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{A}^* = \boldsymbol{B}\boldsymbol{B}^*$ is simply $\boldsymbol{X} = \boldsymbol{H}$. The minimum and maximum eigenvalues of $\boldsymbol{A}$ are $\lambda_{min}(\boldsymbol{A}) = \frac{1}{2}$ and $\lambda_{max}(\boldsymbol{A}) = N - \frac{1}{2}$, and thus the condition number for $\boldsymbol{A}$ is $\kappa = 2N - 1$. Thus, by applying Lemma A.1 with $\delta = \frac{\delta_H}{\pi}$, we can construct an $N \times R_H$ matrix $\boldsymbol{Z}$ with

$$R_H = \left\lceil \frac{1}{\pi^2}\log(8N-4)\log\left(\frac{4\pi}{\delta_H}\right)\right\rceil$$

such that

$$\|\boldsymbol{H} - \boldsymbol{Z}\boldsymbol{Z}^*\| \leq \frac{\delta_H}{\pi}\|\boldsymbol{H}\|.$$

It is shown in [109] that the operator norm of the infinite Hilbert matrix is bounded above by π, and thus, the finite dimensional matrix $\boldsymbol{H}$ satisfies $\|\boldsymbol{H}\| \leq \pi$. Therefore, $\|\boldsymbol{H} - \boldsymbol{Z}\boldsymbol{Z}^*\| \leq \delta_H$, as desired.

Controlling the rank of $\boldsymbol{A}_1$

Next, we construct a low-rank matrix $\widetilde{\boldsymbol{A}}_1$ such that $\|\boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1\| \leq \delta_A$ for some desired $\delta_A > 0$. In this case we will require a different approach. We begin by noting that by using the Taylor series expansions [18]

$$\frac{1}{\sin \pi x} - \frac{1}{\pi x} = \frac{2}{\pi}\sum_{k=1}^{\infty}(1 - 2^{-(2k-1)})\zeta(2k)x^{2k-1},$$

and

$$\frac{1}{\pi(x+1)} + \frac{1}{\pi(x-1)} = -\frac{2x}{\pi(1-x^2)} = -\frac{2}{\pi}\sum_{k=1}^{\infty} x^{2k-1},$$

we can write

$$\begin{aligned}\boldsymbol{A}_1[m,n] &= \frac{1}{\pi(m-n)} - \frac{1}{N\sin\left(\pi\frac{m-n}{N}\right)} - \frac{1}{\pi(m-n+N)} - \frac{1}{\pi(m-n-N)} \\ &= \frac{2}{N\pi}\sum_{k=1}^{\infty}\left[1 - (1-2^{-(2k-1)})\zeta(2k)\right]\left(\frac{m-n}{N}\right)^{2k-1}.\end{aligned}$$

We can then define a new $N \times N$ matrix $\widetilde{\boldsymbol{A}}_1$ by truncating the series to R_A terms:

$$\widetilde{\boldsymbol{A}}_1[m,n] := \frac{2}{N\pi}\sum_{k=1}^{R_A}\left[1 - (1-2^{-(2k-1)})\zeta(2k)\right]\left(\frac{m-n}{N}\right)^{2k-1}.$$

Note that each entry of $\widetilde{\boldsymbol{A}}_1$ is a polynomial of degree $2R_A - 1$ in both m and n. Thus, we could also write

$$\widetilde{\boldsymbol{A}}_1[m,n] = \sum_{k=0}^{2R_A-1}\sum_{\ell=0}^{2R_A-1} c_{k,\ell}m^k n^\ell,$$

[18] Here, $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$ is the Riemann-Zeta function.

for a set of scalars $c_{k,\ell} \in \mathbb{R}$. If we let $\boldsymbol{V}_A$ be the $N \times 2R_A$ matrix with entries $\boldsymbol{V}_A[m,k] = m^k$ and let $\boldsymbol{C}_A$ be the $2R_A \times 2R_A$ matrix with entries $\boldsymbol{C}_A[k,\ell] = c_{k,\ell}$, it is easy to see that we can write $\widetilde{\boldsymbol{A}}_1 = \boldsymbol{V}_A \boldsymbol{C}_A \boldsymbol{V}_A^*$. Thus, $\widetilde{\boldsymbol{A}}_1$ has rank $2R_A$.

Next, we note that by using the identity $(1-2^{1-s})\zeta(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^s}$ for $s > 1$, we have

$$0 \leq 1 - (1 - 2^{-(2k-1)})\zeta(2k) = \sum_{n=2}^{\infty} \frac{(-1)^n}{n^{2k}} \leq \frac{1}{2^{2k}},$$

where the inequality follows from the fact that this is an alternating series whose terms decrease in magnitude. Hence, we can bound the truncation error $|\boldsymbol{A}_1[m,n] - \widetilde{\boldsymbol{A}}_1[m,n]|$ by

$$\begin{aligned}
|\boldsymbol{A}_1[m,n] - \widetilde{\boldsymbol{A}}_1[m,n]| &= \left| \frac{2}{N\pi} \sum_{k=R_A+1}^{\infty} \left[1 - (1 - 2^{-(2k-1)})\zeta(2k)\right] \left(\frac{m-n}{N}\right)^{2k-1} \right| \\
&\leq \frac{2}{N\pi} \sum_{k=R_A+1}^{\infty} \left[1 - (1 - 2^{-(2k-1)})\zeta(2k)\right] \left|\frac{m-n}{N}\right|^{2k-1} \\
&\leq \frac{2}{N\pi} \sum_{k=R_A+1}^{\infty} \frac{1}{2^{2k}} \cdot 1 \\
&= \frac{2}{3N\pi} \left(\frac{1}{2}\right)^{2R_A}.
\end{aligned}$$

Therefore, the error $\|\boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1\|_F^2$ is bounded by:

$$\|\boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1\|_F^2 = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} |\boldsymbol{A}_1[m,n] - \widetilde{\boldsymbol{A}}_1[m,n]|^2 \leq N \cdot N \cdot \left[\frac{2}{3N\pi} \left(\frac{1}{2}\right)^{2R_A}\right]^2 = \frac{4}{9\pi^2} \left(\frac{1}{2}\right)^{4R_A}.$$

Hence,

$$\|\boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1\| \leq \|\boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1\|_F \leq \frac{2}{3\pi} \left(\frac{1}{2}\right)^{2R_A}.$$

Thus, for any $\delta_A \in (0, \frac{8}{3\pi})$ we can set $R_A = \left\lceil \frac{1}{2\log 2} \log\left(\frac{2}{3\pi\delta_A}\right) \right\rceil$ and ensure that $\|\boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1\| \leq \delta_A$ and

$$\operatorname{rank}(\widetilde{\boldsymbol{A}}_1) \leq 2 \left\lceil \frac{1}{2\log 2} \log\left(\frac{2}{3\pi\delta_A}\right) \right\rceil.$$

Controlling the rank of $\boldsymbol{B}_0$

To construct a low-rank matrix $\widetilde{\boldsymbol{B}}_0$ such that $\|\boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0\| \leq \delta_B$ for some desired $\delta_B > 0$, we use a similar approach as above. Using the Taylor series

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!},$$

we can write

$$\begin{aligned}
\boldsymbol{B}_0[m,n] &= \frac{2\sin(\pi(W-W')(m-n))}{\pi(m-n)} \\
&= \sum_{k=0}^{\infty} \frac{2(-1)^k[\pi(W-W')(m-n)]^{2k+1}}{(2k+1)!\pi(m-n)} \\
&= \frac{2}{N\pi} \sum_{k=0}^{\infty} \frac{(-1)^k[\pi(W-W')N]^{2k+1}}{(2k+1)!} \left(\frac{m-n}{N}\right)^{2k}.
\end{aligned}$$

We can then define a new matrix $\widetilde{\boldsymbol{B}}_0$ by truncating the series to R_B terms:

$$\widetilde{\boldsymbol{B}}_0[m,n] := \frac{2}{N\pi} \sum_{k=0}^{R_B-1} \frac{(-1)^k[\pi(W-W')N]^{2k+1}}{(2k+1)!} \left(\frac{m-n}{N}\right)^{2k}.$$

Note that each entry of $\widetilde{\boldsymbol{B}}_0$ is a polynomial of degree $2R_B - 2$ in both m and n. Thus, we could also write

$$\widetilde{\boldsymbol{B}}_0[m,n] = \sum_{k=0}^{2R_B-2} \sum_{\ell=0}^{2R_B-2} c'_{k,\ell} m^k n^\ell,$$

for a set of scalars $c'_{k,\ell} \in \mathbb{R}$. If we let $\boldsymbol{V}_B$ be the $N \times (2R_B - 1)$ matrix with entries $\boldsymbol{V}_B[m,k] = m^k$ and let $\boldsymbol{C}_B$ be the $(2R_B-1) \times (2R_B-1)$ matrix with entries $\boldsymbol{C}_B[k,\ell] = c'_{k,\ell}$, it is easy to see that we can write $\widetilde{\boldsymbol{B}}_0 = \boldsymbol{V}_B \boldsymbol{C}_B \boldsymbol{V}_B^*$. Thus, $\widetilde{\boldsymbol{B}}_0$ has rank $2R_B - 1$.

Next, we note that by definition, $2W'N$ is $2WN$ rounded to the nearest odd integer. Hence, $|2W'N - 2WN| \leq 1$, and so, $|\pi(W-W')N| \leq \frac{\pi}{2}$. Also, we have that $(2k+1)! \geq \frac{2}{9} \cdot 3^{2k+1}$ for all integers $k \geq 0$. Hence, the truncation error $|\boldsymbol{B}_0[m,n] - \widetilde{\boldsymbol{B}}_0[m,n]|$ is bounded by:

$$
\begin{aligned}
|\boldsymbol{B}_0[m,n] - \widetilde{\boldsymbol{B}}_0[m,n]| &\leq \left| \frac{2}{N\pi} \sum_{k=R_B}^{\infty} \frac{(-1)^k [\pi(W-W')N]^{2k+1}}{(2k+1)!} \left(\frac{m-n}{N}\right)^{2k} \right| \\
&\leq \frac{2}{N\pi} \left| \frac{(-1)^{R_B} [\pi(W-W')N]^{2r_b+1}}{(2K+1)!} \left(\frac{m-n}{N}\right)^{2K} \right| \\
&\leq \frac{2}{N\pi} \frac{\left(\frac{\pi}{2}\right)^{2R_B+1}}{\frac{2}{9} \cdot 3^{2R_B+1}} \cdot 1 \\
&= \frac{3}{2N} \left(\frac{\pi}{6}\right)^{2R_B},
\end{aligned}
$$

where we have used the fact that an alternating series whose terms decrease in magnitude can be bounded by the magnitude of the first term. Thus, the error $\|\boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0\|_F^2$ is bounded by:

$$
\|\boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0\|_F^2 = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} |\boldsymbol{B}_0[m,n] - \widetilde{\boldsymbol{B}}_0[m,n]|^2 \leq N \cdot N \cdot \left[\frac{3}{2N} \left(\frac{\pi}{6}\right)^{2R_B} \right]^2 = \frac{9}{4} \left(\frac{\pi}{6}\right)^{4R_B}.
$$

Hence,

$$
\|\boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0\| \leq \|\boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0\|_F \leq \frac{3}{2} \left(\frac{\pi}{6}\right)^{2R_B}.
$$

Thus, for any $\delta_B \in (0, \frac{3}{2})$ we can set $R_B = \left\lceil \frac{1}{2 \log \frac{6}{\pi}} \log\left(\frac{3}{2\delta_B}\right) \right\rceil$ and ensure that $\|\boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0\| \leq \delta_B$ and

$$
\operatorname{rank}(\widetilde{\boldsymbol{B}}_0) \leq 2 \left\lceil \frac{1}{2 \log \frac{6}{\pi}} \log\left(\frac{3}{2\delta_B}\right) \right\rceil - 1.
$$

Putting it all together

For any $\epsilon \in (0, \frac{1}{2})$, set[19] $\delta_H = \frac{4\pi}{15}\epsilon$ and $\delta_A = \delta_B = \frac{7}{30}\epsilon$. Then, let $\widetilde{\boldsymbol{H}} = \boldsymbol{Z}\boldsymbol{Z}^*$, $\widetilde{\boldsymbol{A}}_1 = \boldsymbol{V}_A \boldsymbol{C}_A \boldsymbol{V}_A^*$, and $\widetilde{\boldsymbol{B}}_0 = \boldsymbol{V}_B \boldsymbol{C}_B \boldsymbol{V}_B^*$ be defined as in the previous subsections. Also, define $\boldsymbol{E}_H = \boldsymbol{H} - \widetilde{\boldsymbol{H}}$, $\boldsymbol{E}_A = \boldsymbol{A}_1 - \widetilde{\boldsymbol{A}}_1$, $\boldsymbol{E}_B = \boldsymbol{B}_0 - \widetilde{\boldsymbol{B}}_0$. By using these definitions along with (A.3), we can write

$$
\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W} \boldsymbol{F}_{N,W}^* = \boldsymbol{L} + \boldsymbol{E}_F,
$$

[19] It may be possible to obtain a slightly better bound via a more careful selection of δ_A, δ_B, and δ_H. We have not pursued such refinements here as there is not much room for significant improvement.

where

$$\begin{aligned}\boldsymbol{L} &= \tfrac{1}{2j}\boldsymbol{D}_A\left[\tfrac{1}{\pi}(\widetilde{\boldsymbol{H}}\boldsymbol{J}-\boldsymbol{J}\widetilde{\boldsymbol{H}})+\widetilde{\boldsymbol{A}}_1\right]\boldsymbol{D}_A^* - \tfrac{1}{2j}\boldsymbol{D}_A^*\left[\tfrac{1}{\pi}(\widetilde{\boldsymbol{H}}\boldsymbol{J}-\boldsymbol{J}\widetilde{\boldsymbol{H}})+\widetilde{\boldsymbol{A}}_1\right]\boldsymbol{D}_A \\ &\quad + \tfrac{1}{2}\boldsymbol{D}_B\widetilde{\boldsymbol{B}}_0\boldsymbol{D}_B^* + \tfrac{1}{2}\boldsymbol{D}_B^*\widetilde{\boldsymbol{B}}_0\boldsymbol{D}_B \\ &= \tfrac{1}{2\pi j}(\boldsymbol{D}_A\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{J}\boldsymbol{D}_A^* - \boldsymbol{D}_A\boldsymbol{J}\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{D}_A^* - \boldsymbol{D}_A^*\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{J}\boldsymbol{D}_A - \boldsymbol{D}_A^*\boldsymbol{J}\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{D}_A) \\ &\quad + \tfrac{1}{2j}(\boldsymbol{D}_A\boldsymbol{V}_A\boldsymbol{C}_A\boldsymbol{V}_A^*\boldsymbol{D}_A^* - \boldsymbol{D}_A^*\boldsymbol{V}_A\boldsymbol{C}_A\boldsymbol{V}_A^*\boldsymbol{D}_A) + \tfrac{1}{2}(\boldsymbol{D}_B\boldsymbol{V}_B\boldsymbol{C}_B\boldsymbol{V}_B^*\boldsymbol{D}_B^* + \boldsymbol{D}_B^*\boldsymbol{V}_B\boldsymbol{C}_B\boldsymbol{V}_B^*\boldsymbol{D}_B)\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{E}_F =& \tfrac{1}{2j}\boldsymbol{D}_A\left[\tfrac{1}{\pi}(\boldsymbol{E}_H\boldsymbol{J}-\boldsymbol{J}\boldsymbol{E}_H)+\boldsymbol{E}_A\right]\boldsymbol{D}_A^* - \tfrac{1}{2j}\boldsymbol{D}_A^*\left[\tfrac{1}{\pi}(\boldsymbol{E}_H\boldsymbol{J}-\boldsymbol{J}\boldsymbol{E}_H)+\boldsymbol{E}_A\right]\boldsymbol{D}_A \\ &+ \tfrac{1}{2}\boldsymbol{D}_B\boldsymbol{E}_B\boldsymbol{D}_B^* + \tfrac{1}{2}\boldsymbol{D}_B^*\boldsymbol{E}_B\boldsymbol{D}_B.\end{aligned}$$

If we define

$$\begin{aligned}&\boldsymbol{L}_1 = \\ &\left[\tfrac{1}{2\pi j}\boldsymbol{D}_A\boldsymbol{Z} \quad -\tfrac{1}{2\pi j}\boldsymbol{D}_A\boldsymbol{J}\boldsymbol{Z} \quad -\tfrac{1}{2\pi j}\boldsymbol{D}_A^*\boldsymbol{Z} \quad \tfrac{1}{2\pi j}\boldsymbol{D}_A^*\boldsymbol{J}\boldsymbol{Z} \quad \tfrac{1}{2j}\boldsymbol{D}_A\boldsymbol{V}_A \quad -\tfrac{1}{2j}\boldsymbol{D}_A^*\boldsymbol{V}_A \quad \tfrac{1}{2}\boldsymbol{D}_B\boldsymbol{V}_B \quad \tfrac{1}{2}\boldsymbol{D}_B^*\boldsymbol{V}_B\right]\end{aligned}$$

and

$$\boldsymbol{L}_2 = \left[\boldsymbol{D}_A\boldsymbol{J}\boldsymbol{Z} \quad \boldsymbol{D}_A\boldsymbol{Z} \quad \boldsymbol{D}_A^*\boldsymbol{J}\boldsymbol{Z} \quad \boldsymbol{D}_A^*\boldsymbol{Z} \quad \boldsymbol{D}_A\boldsymbol{V}_A\boldsymbol{C}_A^* \quad \boldsymbol{D}_A^*\boldsymbol{V}_A\boldsymbol{C}_A^* \quad \boldsymbol{D}_B\boldsymbol{V}_B\boldsymbol{C}_B^* \quad \boldsymbol{D}_B^*\boldsymbol{V}_B\boldsymbol{C}_B^*\right],$$

then $\boldsymbol{L} = \boldsymbol{L}_1\boldsymbol{L}_2^*$ and $\boldsymbol{L}_1, \boldsymbol{L}_2$ are both $N \times R_L$ matrices, where

$$\begin{aligned}R_L &= 4\cdot R_H + 2\cdot 2R_A + 2\cdot(2R_B-1) \\ &= 4\left\lceil\frac{1}{\pi^2}\log(8N-4)\log\left(\frac{4\pi}{\delta_H}\right)\right\rceil + 4\left\lceil\frac{1}{2\log 2}\log\left(\frac{2}{3\pi\delta_A}\right)\right\rceil + 4\left\lceil\frac{1}{2\log\frac{6}{\pi}}\log\left(\frac{3}{2\delta_B}\right)\right\rceil - 2 \\ &\le \frac{4}{\pi^2}\log(8N-4)\log\left(\frac{4\pi}{\delta_H}\right) + \frac{2}{\log 2}\log\left(\frac{2}{3\pi\delta_A}\right) + \frac{2}{\log\frac{6}{\pi}}\log\left(\frac{3}{2\delta_B}\right) + 10 \\ &= \frac{4}{\pi^2}\log(8N-4)\log\left(\frac{15}{\epsilon}\right) + \frac{2}{\log 2}\log\left(\frac{20}{7\pi\epsilon}\right) + \frac{2}{\log\frac{6}{\pi}}\log\left(\frac{45}{7\epsilon}\right) + 10 \\ &= \left(\frac{4}{\pi^2}\log(8N-4) + \frac{2}{\log 2} + \frac{2}{\log\frac{6}{\pi}}\right)\log\left(\frac{15}{\epsilon}\right) + \frac{2}{\log 2}\log\left(\frac{4}{21\pi}\right) + \frac{2}{\log\frac{6}{\pi}}\log\left(\frac{3}{7}\right) + 10 \\ &\le \left(\frac{4}{\pi^2}\log(8N) + 6\right)\log\left(\frac{15}{\epsilon}\right).\end{aligned}$$

Also, by applying the triangle inequality and using the fact that $\|\boldsymbol{D}_A\| = \|\boldsymbol{D}_B\| = \|\boldsymbol{J}\| = 1$, we see that

$$\|\boldsymbol{E}_F\| \leq \frac{2}{\pi}\|\boldsymbol{E}_H\| + \|\boldsymbol{E}_A\| + \|\boldsymbol{E}_B\| \leq \frac{2}{\pi}\delta_H + \delta_A + \delta_B = \frac{2}{\pi} \cdot \frac{4\pi\epsilon}{15} + \frac{7\epsilon}{30} + \frac{7\epsilon}{30} = \epsilon.$$

Together, these two facts establish the theorem.

A.2 Proof of Corollary 3.1

Corollary 3.1 is a direct consequence of Theorem 3.1 together with the following lemma.

Lemma A.2. *Let $\boldsymbol{A}$ be an $N \times N$ Hermitian matrix with eigenvalues $\lambda^{(0)} \geq \cdots \geq \lambda^{(N-1)}$. Suppose we can write*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{U}^* + \boldsymbol{L} + \boldsymbol{E},$$

where $\boldsymbol{U}$ is an $N \times K$ matrix with orthonormal columns ($\boldsymbol{U}^*\boldsymbol{U} = \mathbf{I}$), $\boldsymbol{L}$ is an $N \times N$ Hermitian matrix with $\operatorname{rank}(\boldsymbol{L}) = r$, and $\boldsymbol{E}$ is an $N \times N$ Hermitian matrix with $\|\boldsymbol{E}\| \leq \epsilon$. Then

$$\#\{\ell \ : \ \epsilon < \lambda^{(\ell)} < 1 - \epsilon\} \ \leq \ 2r.$$

Proof. Define $\mathbb{S}_1 = \operatorname{Null}(\boldsymbol{U}\boldsymbol{U}^*) \cap \operatorname{Null}(\boldsymbol{L})$ and $d_1 = \dim(\mathbb{S}_1)$. For any $\boldsymbol{x} \in \mathbb{S}_1$ with $\|\boldsymbol{x}\|_2 = 1$,

$$\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^*\boldsymbol{E}\boldsymbol{x} \leq \|\boldsymbol{E}\| \leq \epsilon.$$

Then by the Courant-Fischer-Weyl min-max theorem,

$$\lambda^{(N-d_1)} = \min_{\substack{\text{subspaces } \mathbb{S} \\ \dim(\mathbb{S})=d_1}} \left[\max_{\substack{\boldsymbol{x}\in\mathbb{S} \\ \|\boldsymbol{x}\|_2=1}} \boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} \right] \ \leq \ \max_{\substack{\boldsymbol{x}\in\mathbb{S}_1 \\ \|\boldsymbol{x}\|_2=1}} \boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} \leq \epsilon.$$

Similarly, let $\mathbb{S}_2 = \operatorname{Null}(\mathbf{I} - \boldsymbol{U}\boldsymbol{U}^*) \cap \operatorname{Null}(\boldsymbol{L})$ and $d_2 = \dim(\mathbb{S}_2)$. Then, for any $\boldsymbol{x} \in \mathbb{S}_2$ with $\|\boldsymbol{x}\|_2 = 1$,

$$\boldsymbol{x}^*(\mathbf{I} - \boldsymbol{A})\boldsymbol{x} = -\boldsymbol{x}^*\boldsymbol{E}\boldsymbol{x} \leq \|\boldsymbol{E}\| \leq \epsilon.$$

Since the eigenvalues of $\mathbf{I} - \boldsymbol{A}$ are $1 - \lambda^{(N-1)} \geq \cdots \geq 1 - \lambda^{(0)}$, the min-max theorem tells us

$$1 - \lambda^{(d_2-1)} = \min_{\substack{\text{subspaces } \mathbb{S} \\ \dim(\mathbb{S})=d_2}} \left[\max_{\substack{\boldsymbol{x}\in\mathbb{S} \\ \|\boldsymbol{x}\|_2=1}} \boldsymbol{x}^*(\mathbf{I} - \boldsymbol{A})\boldsymbol{x} \right] \ \leq \ \max_{\substack{\boldsymbol{x}\in\mathbb{S}_2 \\ \|\boldsymbol{x}\|_2=1}} \boldsymbol{x}^*(\mathbf{I} - \boldsymbol{A})\boldsymbol{x} \leq \epsilon,$$

meaning $\lambda^{(d_2-1)} \geq 1-\epsilon$. Thus

$$\#\{\ell \ : \ \epsilon < \lambda^{(\ell)} < 1-\epsilon\} \leq N - d_1 - d_2.$$

Since for any two $N \times N$ matrices $\boldsymbol{M}_1, \boldsymbol{M}_2$

$$\dim(\mathrm{Null}(\boldsymbol{M}_1) \cap \mathrm{Null}(\boldsymbol{M}_2)) \geq N - (\dim(\mathrm{Range}(\boldsymbol{M}_1)) + \dim(\mathrm{Range}(\boldsymbol{M}_2))),$$

we know $d_1 \geq N - (K + r)$ and $d_2 \geq N - (N - K + r)$. Thus,

$$\#\{\ell \ : \ \epsilon < \lambda^{(\ell)} < 1-\epsilon\} \ \leq \ 2r.$$

□

A.3 Proof of Corollary 3.2

Corollary 3.2 is a direct consequence of Corollary 3.1 together with the following lemma.

Lemma A.3. *Let $\boldsymbol{A}$ be an $N \times N$ symmetric matrix with eigenvalues $1 \geq \lambda^{(0)} \geq \cdots \geq \lambda^{(N-1)} \geq 0$ and corresponding eigenvectors $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{N-1}$. Fix $\epsilon \in (0, \frac{1}{2})$, and let*

$$r' = \#\{\ell \ : \ \epsilon < \lambda^{(\ell)} < 1-\epsilon\}.$$

Choose K such that $\lambda^{(K-1)} > \epsilon$ and $\lambda^{(K)} < 1-\epsilon$, and set $\boldsymbol{V}_{[K]} = \begin{bmatrix} \boldsymbol{v}_0 & \cdots & \boldsymbol{v}_{K-1} \end{bmatrix}$. Then there exists $N \times r'$ matrices $\boldsymbol{U}_1, \boldsymbol{U}_2$ and an $N \times N$ matrix $\boldsymbol{E}$ with $\|\boldsymbol{E}\| \leq \epsilon$ such that

$$\boldsymbol{V}_{[K]}\boldsymbol{V}_{[K]}^* = \boldsymbol{A} + \boldsymbol{U}_1\boldsymbol{U}_2^* + \boldsymbol{E}.$$

Proof. First, we partition the eigenvalues of $\boldsymbol{A}$ into four sets:

$$\mathbb{I}_1 = \{\ell : \lambda^{(\ell)} \geq 1-\epsilon\}, \quad \mathbb{I}_2 = \{\ell : \ell < K, \ \epsilon < \lambda^{(\ell)} < 1-\epsilon\},$$
$$\mathbb{I}_3 = \{\ell : \ell \geq K, \ \epsilon < \lambda^{(\ell)} < 1-\epsilon\}, \quad \mathbb{I}_4 = \{\ell : \lambda^{(\ell)} \leq \epsilon\}.$$

We can write

$$\boldsymbol{A} = \boldsymbol{V}_1\boldsymbol{\Lambda}_1\boldsymbol{V}_1^* + \boldsymbol{V}_2\boldsymbol{\Lambda}_2\boldsymbol{V}_2^* + \boldsymbol{V}_3\boldsymbol{\Lambda}_3\boldsymbol{V}_3^* + \boldsymbol{V}_4\boldsymbol{\Lambda}_4\boldsymbol{V}_4^*$$

and

$$\boldsymbol{V}_{[K]}\boldsymbol{V}_{[K]}^* = \boldsymbol{V}_1\boldsymbol{V}_1^* + \boldsymbol{V}_2\boldsymbol{V}_2^*,$$

where the $\boldsymbol{V}_i$ contain the eigenvectors from $\mathbb{I}_i$ as their columns, and the $\boldsymbol{\Lambda}_i$ are diagonal containing the corresponding eigenvalues. Thus,

$$\begin{aligned}\boldsymbol{V}_{[K]}\boldsymbol{V}_{[K]}^* - \boldsymbol{A} &= [\boldsymbol{V}_2(\mathbf{I}-\boldsymbol{\Lambda}_2)\boldsymbol{V}_2^* - \boldsymbol{V}_3\boldsymbol{\Lambda}_3\boldsymbol{V}_3^*] + [\boldsymbol{V}_1(\mathbf{I}-\boldsymbol{\Lambda}_1)\boldsymbol{V}_1^* - \boldsymbol{V}_4\boldsymbol{\Lambda}_4\boldsymbol{V}_4^*] \\ &=: \boldsymbol{U}_1\boldsymbol{U}_2^* + \boldsymbol{E},\end{aligned}$$

where

$$\boldsymbol{U}_1 = \begin{bmatrix}\boldsymbol{V}_2(\mathbf{I}-\boldsymbol{\Lambda}_2)^{1/2} & -\boldsymbol{V}_3\boldsymbol{\Lambda}_3^{1/2}\end{bmatrix},$$

$$\boldsymbol{U}_2 = \begin{bmatrix}\boldsymbol{V}_2(\mathbf{I}-\boldsymbol{\Lambda}_2)^{1/2} & \boldsymbol{V}_3\boldsymbol{\Lambda}_3^{1/2}\end{bmatrix},$$

$$\boldsymbol{E} = \begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_4\end{bmatrix}\begin{bmatrix}\mathbf{I}-\boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\Lambda}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^* \\ \boldsymbol{V}_4^*\end{bmatrix}.$$

Notice that the number of columns in both $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ is the same as the size of $\mathbb{I}_2 \cup \mathbb{I}_3$, which is exactly r'. Also, since both $\|\mathbf{I}-\boldsymbol{\Lambda}_1\| \leq \epsilon$ and $\|\boldsymbol{\Lambda}_4\| \leq \epsilon$, and $\begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_4\end{bmatrix}$ has orthonormal columns, we have $\|\boldsymbol{E}\| \leq \epsilon$. □

A.4 Proof of Corollary 3.3

Corollary 3.3 is a direct consequence of Corollary 3.1 together with the following lemma.

Lemma A.4. *Let $\boldsymbol{A}$ be an $N \times N$ symmetric matrix with eigenvalues $1 \geq \lambda^{(0)} \geq \cdots \geq \lambda^{(N-1)} \geq 0$ and corresponding eigenvectors $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{N-1}$. Fix $\epsilon \in (0, \frac{1}{2})$, and let*

$$r = \#\{\ell \ : \ \epsilon < \lambda^{(\ell)} < 1-\epsilon\}.$$

Choose K such that $\lambda^{(K-1)} > \epsilon$ and $\lambda^{(K)} < 1-\epsilon$. Let $\boldsymbol{V}_{[K]} = \begin{bmatrix}\boldsymbol{v}_0 & \cdots & \boldsymbol{v}_{K-1}\end{bmatrix}$ and $\boldsymbol{\Lambda}_{[K]} = \mathrm{diag}(\lambda^{(0)}, \ldots, \lambda^{(K-1)})$. Define $\boldsymbol{A}_K^\dagger = \boldsymbol{V}_{[K]}\boldsymbol{\Lambda}_{[K]}^{-1}\boldsymbol{V}_{[K]}$ to be the rank-K truncated pseudoinverse of $\boldsymbol{A}$. Then there exists $N \times r$ matrices $\boldsymbol{U}_3, \boldsymbol{U}_4$ and an $N \times N$ matrix $\boldsymbol{E}$ with $\|\boldsymbol{E}\| \leq 3\epsilon$ such that

$$\boldsymbol{A}_K^\dagger = \boldsymbol{A} + \boldsymbol{U}_3\boldsymbol{U}_4^* + \boldsymbol{E}.$$

Proof. We partition the eigenvalues of $\boldsymbol{A}$ into four sets:

$$\begin{aligned}\mathbb{I}_1 &= \{\ell : \lambda^{(\ell)} \geq 1-\epsilon\}, \quad \mathbb{I}_2 = \{\ell : \ell < K,\ \epsilon < \lambda^{(\ell)} < 1-\epsilon\},\\ \mathbb{I}_3 &= \{\ell : \ell \geq K,\ \epsilon < \lambda^{(\ell)} < 1-\epsilon\}, \quad \mathbb{I}_4 = \{\ell : \lambda^{(\ell)} \leq \epsilon\}.\end{aligned}$$

We can write

$$\boldsymbol{A} = \boldsymbol{V}_1\boldsymbol{\Lambda}_1\boldsymbol{V}_1^* + \boldsymbol{V}_2\boldsymbol{\Lambda}_2\boldsymbol{V}_2^* + \boldsymbol{V}_3\boldsymbol{\Lambda}_3\boldsymbol{V}_3^* + \boldsymbol{V}_4\boldsymbol{\Lambda}_4\boldsymbol{V}_4^*,$$

and

$$\boldsymbol{A}_K^\dagger = \boldsymbol{V}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{V}_1^* + \boldsymbol{V}_2\boldsymbol{\Lambda}_2^{-1}\boldsymbol{V}_2^*$$

where the $\boldsymbol{V}_i$ contain the eigenvectors from $\mathbb{I}_i$ as their columns, and the $\boldsymbol{\Lambda}_i$ are diagonal containing the corresponding eigenvalues. Thus,

$$\begin{aligned}\boldsymbol{A}_K^\dagger - \boldsymbol{A} &= \left[\boldsymbol{V}_2(\boldsymbol{\Lambda}_2^{-1} - \boldsymbol{\Lambda}_2)\boldsymbol{V}_2^* - \boldsymbol{V}_3\boldsymbol{\Lambda}_3\boldsymbol{V}_3^*\right] + \left[\boldsymbol{V}_1(\boldsymbol{\Lambda}_1^{-1} - \boldsymbol{\Lambda}_1)\boldsymbol{V}_1^* - \boldsymbol{V}_4\boldsymbol{\Lambda}_4\boldsymbol{V}_4^*\right]\\ &=: \boldsymbol{U}_3\boldsymbol{U}_4^* + \boldsymbol{E},\end{aligned}$$

where

$$\boldsymbol{U}_3 = \begin{bmatrix}\boldsymbol{V}_2(\boldsymbol{\Lambda}_2^{-1} - \boldsymbol{\Lambda}_2)^{1/2} & -\boldsymbol{V}_3\boldsymbol{\Lambda}_3^{1/2}\end{bmatrix},$$

$$\boldsymbol{U}_4 = \begin{bmatrix}\boldsymbol{V}_2(\boldsymbol{\Lambda}_2^{-1} - \boldsymbol{\Lambda}_2)^{1/2} & \boldsymbol{V}_3\boldsymbol{\Lambda}_3^{1/2}\end{bmatrix},$$

$$\boldsymbol{E} = \begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_1^{-1} - \boldsymbol{\Lambda}_1 & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{\Lambda}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^*\\ \boldsymbol{V}_4^*\end{bmatrix}.$$

Notice that the number of columns in both $\boldsymbol{U}_3$ and $\boldsymbol{U}_4$ is the same as the size of $\mathbb{I}_2 \cup \mathbb{I}_3$, which is exactly r. Also, since $\|\boldsymbol{\Lambda}_1^{-1} - \boldsymbol{\Lambda}_1\| \leq \frac{1}{1-\epsilon} - (1-\epsilon) = \frac{\epsilon}{1-\epsilon} + \epsilon \leq 3\epsilon$ and $\|\boldsymbol{\Lambda}_4\| \leq \epsilon$, and $\begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_4\end{bmatrix}$ has orthonormal columns, we have $\|\boldsymbol{E}\| \leq 3\epsilon$. □

A.5 Proof of Corollary 3.4

Corollary 3.4 is a direct consequence of Corollary 3.1 together with the following lemma.

Lemma A.5. *Let $\boldsymbol{A}$ be an $N \times N$ symmetric matrix with eigenvalues $1 \geq \lambda_0 \geq \cdots \geq \lambda_{N-1} \geq 0$ and corresponding eigenvectors $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{N-1}$. For a given regularization parameter $\alpha > 0$, define $\boldsymbol{A}_{tik} = (\boldsymbol{A}^*\boldsymbol{A} + \alpha\mathbf{I})^{-1}\boldsymbol{A}^*$. Fix $\epsilon \in (0, \frac{1}{2})$ and let*

$$r = \#\{\ell : \alpha(1+\alpha)\epsilon < \lambda_\ell < 1 - \tfrac{1}{3}\epsilon\}.$$

Then there exists an $N \times r$ matrix $\boldsymbol{U}_5$ and an $N \times N$ matrix $\boldsymbol{E}$ with $\|\boldsymbol{E}\| \leq \epsilon$ such that

$$\boldsymbol{A}_{tik} = \frac{1}{1+\alpha}\boldsymbol{A} + \boldsymbol{U}_5\boldsymbol{U}_5^* + \boldsymbol{E}.$$

Proof. We partition the eigenvalues of $\boldsymbol{A}$ into two sets:

$$\mathbb{I}_1 = \{\ell : \alpha(1+\alpha)\epsilon < \lambda_\ell < 1 - \tfrac{1}{3}\epsilon\}, \quad \mathbb{I}_2 = \{\ell : \lambda_\ell \leq \alpha(1+\alpha)\epsilon \text{ or } \lambda_\ell \geq 1 - \tfrac{1}{3}\epsilon\}.$$

We can write

$$\boldsymbol{A} = \boldsymbol{V}_1\boldsymbol{\Lambda}_1\boldsymbol{V}_1^* + \boldsymbol{V}_2\boldsymbol{\Lambda}_2\boldsymbol{V}_2^*$$

and

$$\boldsymbol{A}_{\text{tik}} = \boldsymbol{V}_1(\boldsymbol{\Lambda}_1^2 + \alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}_1\boldsymbol{V}_1^* + \boldsymbol{V}_2(\boldsymbol{\Lambda}_2^2 + \alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}_2\boldsymbol{V}_2^*,$$

where the $\boldsymbol{V}_i$ contain the eigenvectors from $\mathbb{I}_i$ as their columns, and the $\boldsymbol{\Lambda}_i$ are diagonal containing the corresponding eigenvalues. Thus,

$$\begin{aligned}\boldsymbol{A}_{\text{tik}} - \tfrac{1}{1+\alpha}\boldsymbol{A} &= \boldsymbol{V}_1\left[(\boldsymbol{\Lambda}_1^2 + \alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}_1 - \tfrac{1}{1+\alpha}\boldsymbol{\Lambda}_1\right]\boldsymbol{V}_1^* + \boldsymbol{V}_2\left[(\boldsymbol{\Lambda}_2^2 + \alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}_2 - \tfrac{1}{1+\alpha}\boldsymbol{\Lambda}_2\right]\boldsymbol{V}_2^* \\ &=: \boldsymbol{U}_5\boldsymbol{U}_5^* + \boldsymbol{E},\end{aligned}$$

where

$$\boldsymbol{U}_5 = \boldsymbol{V}_1\left[(\boldsymbol{\Lambda}_1^2 + \alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}_1 - \tfrac{1}{1+\alpha}\boldsymbol{\Lambda}_1\right]^{1/2},$$

$$\boldsymbol{E} = \boldsymbol{V}_2 \left[(\boldsymbol{\Lambda}_2^2 + \alpha \mathbf{I})^{-1} \boldsymbol{\Lambda}_2 - \tfrac{1}{1+\alpha} \boldsymbol{\Lambda}_2 \right] \boldsymbol{V}_2^*.$$

Notice that the number of columns in $\boldsymbol{U}_5$ is the same as the size of $\mathbb{I}_1$, which is exactly r.

Observe that the matrix $(\boldsymbol{\Lambda}_2^2 + \alpha \mathbf{I})^{-1} \boldsymbol{\Lambda}_2 - \frac{1}{1+\alpha} \boldsymbol{\Lambda}_2$ is diagonal, and the diagonal entries are of the form $\frac{\lambda_\ell}{\lambda_\ell^2+\alpha} - \frac{\lambda_\ell}{1+\alpha} = \frac{\lambda_\ell(1-\lambda_\ell^2)}{(1+\alpha)(\lambda_\ell^2+\alpha)}$ where λ_ℓ satisfies either $0 \leq \lambda_\ell \leq \alpha(1+\alpha)\epsilon$ or $1 - \frac{1}{3}\epsilon \leq \lambda_\ell \leq 1$. If $\lambda_\ell \leq \alpha(1+\alpha)\epsilon$, then we have:

$$0 \leq \frac{\lambda_\ell(1-\lambda_\ell^2)}{(1+\alpha)(\lambda_\ell^2+\alpha)} \leq \frac{\alpha(1+\alpha)\epsilon \cdot 1}{(1+\alpha)\alpha} = \epsilon.$$

If $1 - \frac{1}{3}\epsilon \leq \lambda_\ell \leq 1$, then since $0 < \epsilon < \frac{1}{2}$, we also have $\lambda_\ell \geq 1 - \frac{1}{3}\epsilon \geq \frac{5}{6}$, and thus:

$$0 \leq \frac{\lambda_\ell(1-\lambda_\ell^2)}{(1+\alpha)(\lambda_\ell^2+\alpha)} = \frac{\lambda_\ell(1+\lambda_\ell)(1-\lambda_\ell)}{(1+\alpha)(\lambda_\ell^2+\alpha)} \leq \frac{1 \cdot 2 \cdot \frac{1}{3}\epsilon}{1 \cdot (\frac{5}{6})^2} = \frac{24}{25}\epsilon \leq \epsilon.$$

In either case, $0 \leq \frac{\lambda_\ell(1-\lambda_\ell^2)}{(1+\alpha)(\lambda_\ell^2+\alpha)} \leq \epsilon$. Hence, $\|(\boldsymbol{\Lambda}_2^2 + \alpha \mathbf{I})^{-1} \boldsymbol{\Lambda}_2 - \frac{1}{1+\alpha} \boldsymbol{\Lambda}_2\| \leq \epsilon$, and thus $\|\boldsymbol{E}\| \leq \epsilon$. □

A.6 Proof of Lemma A.1

Iterative methods for efficiently computing a low-rank approximation to the solution of a Lyapunov system have been well-studied [90, 134]. The CF-ADI algorithm presented in [88] constructs a factor $\boldsymbol{Z}$ by concatenating a series of r $N \times M$ matrices, $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \cdots & \boldsymbol{Z}_r \end{bmatrix}$, where

$$\begin{aligned} \boldsymbol{Z}_1 &= \sqrt{2p_1}(\boldsymbol{A} + p_1 \mathbf{I})^{-1} \boldsymbol{B} \\ \boldsymbol{Z}_k &= \sqrt{\frac{p_k}{p_{k-1}}} \left(\mathbf{I} - (p_k + p_{k-1})(\boldsymbol{A} + p_k \mathbf{I})^{-1} \right) \boldsymbol{Z}_{k-1}, \quad k = 2, \ldots, r, \end{aligned}$$

for some choice of positive real numbers $p_1, \ldots, p_r$. They show that the matrix $\boldsymbol{Z}\boldsymbol{Z}^*$ produced by this iteration is equivalant to the matrix produced by the ADI iteration given in [90], and thus, $\boldsymbol{Z}\boldsymbol{Z}^*$ satisfies

$$\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{Z}^* = \phi(\boldsymbol{A})\boldsymbol{X}\phi(\boldsymbol{A})^* \text{ where } \phi(x) = \prod_{j=1}^{r} \frac{x - p_j}{x + p_j}.$$

(This is shown in [90] by using induction on r.) Therefore, the error $\|\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{Z}^*\|$ satisfies

$$\|\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{Z}^*\| \leq \|\boldsymbol{X}\| \cdot \|\phi(\boldsymbol{A})\|^2 = \|\boldsymbol{X}\| \cdot \max_{x \in \text{Spec}(\boldsymbol{A})} |\phi(x)|^2 \leq \|\boldsymbol{X}\| \cdot \max_{x \in [a,b]} |\phi(x)|^2,$$

where $a = \lambda_{\min}(\boldsymbol{A})$ and $b = \lambda_{\max}(\boldsymbol{A})$ (so $\kappa = \frac{b}{a}$). In [134], it is shown that for a given interval $[a, b]$ and a number of ADI iterations r, there exists a choice of parameters $p_1, \ldots, p_r$ such that $\max_{x \in [a,b]} |\phi(x)|^2 = \alpha$, where α satisfies

$$\frac{I(\sqrt{1 - \alpha^2})}{I(\alpha)} = \frac{4rI(\kappa^{-1})}{I(\sqrt{1 - \kappa^{-2}})},$$

where $I(\tau)$ is the complete elliptic integral of the first kind, defined by

$$I(\tau) := \int_0^{\pi/2} (1 - \tau^2 \sin^2 \theta)^{-1/2} \, d\theta.$$

It is shown in [85] that the elliptic nome, defined as

$$q(\tau) := \exp\left[-\pi \frac{I(\sqrt{1 - \tau^2})}{I(\tau)}\right]$$

satisfies

$$\tau^2 = 16q(\tau) \prod_{n=1}^{\infty} \left(\frac{1 + q(\tau)^{2n}}{1 + q(\tau)^{2n-1}}\right)^8.$$

For $0 \leq \tau \leq 1$, the range of the elliptic nome is $0 \leq q(\tau) \leq 1$. Hence, the above equation gives us the inequality $\tau^2 \leq 16q(\tau)$. By using the definition of the elliptic nome, this inequality becomes

$$\frac{I(\sqrt{1 - \tau^2})}{I(\tau)} \leq \frac{2}{\pi} \log \frac{4}{\tau} \text{ for } 0 \leq \tau \leq 1.$$

So, by setting the number of iterations as $r = \left\lceil \frac{1}{\pi^2} \log(4\kappa) \log\left(\frac{4}{\delta}\right) \right\rceil$, we have

$$\frac{2}{\pi} \log \frac{4}{\alpha} \geq \frac{I(\sqrt{1 - \alpha^2})}{I(\alpha)} = \frac{4rI(\kappa^{-1})}{I(\sqrt{1 - \kappa^{-2}})} \geq \frac{4 \cdot \frac{1}{\pi^2} \log(4\kappa) \log(\frac{4}{\delta})}{\frac{2}{\pi} \log(4\kappa)} = \frac{2}{\pi} \log \frac{4}{\delta}.$$

Hence, $\max_{x \in [a,b]} |\phi(x)|^2 = \alpha \leq \delta$, and thus, $\|\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{Z}^*\| \leq \delta \|\boldsymbol{X}\|$, as desired.

Remark A.1. It is shown in [90] that (A.4) is a good approximation for the number of iterations needed to get the relative error less than δ, provided that $\kappa \gg 1$. It is shown in [134] that (A.4) is a good approximation provided that $r \geq 3$. Here, we have shown that (A.4) is sufficient to guarantee a strict bound on the error.

Remark A.2. The choice of parameters $p_1, \ldots, p_r$ which minimizes $\max_{x\in[a,b]} |\phi(x)|^2$ is given by the formula $p_k = b\mathrm{dn}\left[\frac{2k-1}{2r} I(\sqrt{1-\kappa^{-2}}), \sqrt{1-\kappa^{-2}}\right]$, where $\mathrm{dn}[z,\tau]$ is the Jacobi elliptic function. This function is defined as $\mathrm{dn}[z,\tau] = \sqrt{1-\tau^2\sin^2\varphi}$, where φ satisfies $\int_0^{\varphi}(1-\tau^2\sin^2\theta)^{-1/2}\,d\theta = z$. If the Jacobi elliptic function dn is not available, a suboptimal choice of parameters $p_1, \ldots, p_r$ is given by $p_k = a^{\frac{2k-1}{2r}} b^{\frac{2r-2k+1}{2r}}$, i.e., we can pick the parameters to be evenly spaced on a log scale.

Remark A.3. If the matrix $\boldsymbol{A}$ is diagonal, each iteration of the CF-ADI algorithm above will take $O(N)$ operations. Hence, the matrix $\boldsymbol{Z}$ can be computed in $O(rN)$ operations.

A.7 Proof of Theorem 3.2

Theorem 3.1 implies that the difference between $\boldsymbol{B}_{N,W}$ and $\boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ is effectively low rank. The main idea is to first show that the difference between the two prolate matrices $\boldsymbol{B}_{N,W}$ and $[\overline{\boldsymbol{B}}_{M,W}]_N$ is also effectively low rank. By using the Taylor series

$$\frac{1}{\sin x} - \frac{1}{x} = \sum_{r=1}^{\infty} \frac{2(1-2^{-(2r-1)})\zeta(2r)}{\pi^{2r}} x^{2r-1},$$

where ζ denotes the Riemann-Zeta function, the (m,n)-th entry of the difference $[\overline{\boldsymbol{B}}_{M,W}]_N - \boldsymbol{B}_{N,W}$ is given by

$$\begin{aligned}
&\left([\overline{\boldsymbol{B}}_{M,W}]_N - \boldsymbol{B}_{N,W}\right)[m,n] \\
&= \frac{\sin\left(2\pi W(m-n)\right)}{M\sin\left((m-n)\pi/M\right)} - \frac{\sin\left(2\pi W(m-n)\right)}{\pi(m-n)} \\
&= \left(\frac{1}{\sin\left(\frac{(m-n)\pi}{M}\right)} - \frac{1}{\frac{(m-n)\pi}{M}}\right)\frac{\sin\left(2\pi W(m-n)\right)}{M} \\
&= \sum_{r=1}^{\infty} t(r; m-n) = \boldsymbol{L}_2[m,n] + \boldsymbol{E}_2[m,n]
\end{aligned}$$

for all $m, n = 0, 1, \ldots, N-1$. Here $t(r;k) := \frac{2}{M\pi}\left(1 - 2^{1-2r}\right)\zeta(2r)\left(\frac{k}{M}\right)^{2r-1}\sin\left(2\pi W k\right)$, and $\boldsymbol{L}_2$ and $\boldsymbol{E}_2$ are $N \times N$ matrices with entries

$$\boldsymbol{L}_2[m,n] = \sum_{r=1}^{R} t(r; m-n), \;\; \boldsymbol{E}_2[m,n] = \sum_{r=R+1}^{\infty} t(r; m-n).$$

Define $\boldsymbol{D} \in \mathbb{R}^{2R \times 2R}$ to have entries

$$\boldsymbol{D}[2r-1-p, p] = \frac{2}{M\pi}\left(1 - 2^{-(2r-1)}\right)\zeta(2r)(-1)^p \binom{2r-1}{p}$$

for $1 \leq r \leq R$ and $0 \leq p \leq 2r-1$, and zeros for the remaining entries. Also define $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{N \times 2R}$ such that

$$\boldsymbol{U}[n,r] = \left(\frac{n}{M}\right)^r \sin\left(2\pi W n\right), \;\; \boldsymbol{V}[n,r] = \left(\frac{n}{M}\right)^r \cos\left(2\pi W n\right)$$

for all $0 \leq r \leq 2R-1$ and $0 \leq n \leq N-1$. The rank of $\boldsymbol{L}_2$ then can be identified by noting that

$$\begin{aligned}
\boldsymbol{L}_2[m,n] &= \sum_{r=1}^{R}\sum_{p=0}^{2r-1} \boldsymbol{D}[2r-1-p,p]\left(\frac{m}{M}\right)^{2r-1-p}\left(\frac{n}{M}\right)^{p} \cdot \\
&\quad \left(\sin\left(2\pi W m\right)\cos\left(2\pi W n\right) - \cos\left(2\pi W m\right)\sin\left(2\pi W n\right)\right) \\
&= \sum_{r=1}^{R}\sum_{p=0}^{2r-1} \boldsymbol{D}[2r-1-p,p]\big(\boldsymbol{U}[m,2r-1-p]\boldsymbol{V}[n,p] \\
&\qquad\qquad - \boldsymbol{V}[m,2r-1-p]\boldsymbol{U}[n,p]\big) \\
&= \left(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^*\right)[m,n] - \left(\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^*\right)[m,n].
\end{aligned}$$

This implies $\boldsymbol{L}_2 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^* - \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^*$ and $\mathrm{rank}(\boldsymbol{L}_2) \leq 4R$.

Also note that $1 - 2^{1-s}\zeta(s) = \eta(s)$ is the Dirichlet eta function satisfying $0 < \eta(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} < 1$ for all $s \geq 1$. We now turn to bound the entries in $\boldsymbol{E}_2$ as

$$|\boldsymbol{E}_2(m,n)| = \left|\sum_{r=R+1}^{\infty} t(r;m-n)\right| \le \sum_{r=R+1}^{\infty} \frac{2}{\pi M}\left(\frac{N}{M}\right)^{2r-1}$$
$$= \frac{2}{\pi N}\frac{\left(\frac{N}{M}\right)^{2R+2}}{1-\left(\frac{N}{M}\right)^2} = \frac{2}{\pi N}\frac{\left(\frac{N}{M}\right)^{2R}}{\left(\frac{M}{N}\right)^2-1}.$$

Choosing $R = \max\left(\frac{-\log\frac{\pi}{32}((\frac{M}{N})^2-1)\epsilon}{2\log\frac{M}{N}}, 0\right)$, we obtain that $|\boldsymbol{E}_2(m,n)| \le \frac{\epsilon}{16N}$. It follows from the Gershgorin circle theorem that

$$\|\boldsymbol{E}_2\| \le \max_m \sum_n |\boldsymbol{E}_2(m,n)| \le \frac{\epsilon}{16}.$$

By Theorem 3.1, there exist $N \times N$ matrices $\boldsymbol{L}_1$ and $\boldsymbol{E}_1$ with

$$\operatorname{rank}(\boldsymbol{L}_1) \le \left(\tfrac{4}{\pi^2}\log(8N)+6\right)\log\left(\tfrac{16}{\epsilon}\right), \quad \|\boldsymbol{E}_1\| \le \tfrac{15}{16}\epsilon$$

such that $\boldsymbol{B}_{N,W} = \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^* + \boldsymbol{L}_1 + \boldsymbol{E}_1$.

Denoting $\boldsymbol{L} = \boldsymbol{L}_1 + \boldsymbol{L}_2$ and $\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2$, we obtain

$$[\overline{\boldsymbol{B}}_{M,W}]_N = \boldsymbol{B}_{N,W} + \boldsymbol{L}_2 + \boldsymbol{E}_2 = \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^* + \boldsymbol{L} + \boldsymbol{E},$$

where

$$\operatorname{rank}(\boldsymbol{L}) \le \operatorname{rank}(\boldsymbol{L}_1) + \operatorname{rank}(\boldsymbol{L}_2)$$
$$\le \left(\tfrac{4}{\pi^2}\log(8N)+6\right)\log\left(\tfrac{16}{\epsilon}\right) + 2\max\left(\frac{-\log 8\pi((\frac{M}{N})^2-1)\epsilon}{\log(\frac{M}{N})}, 0\right)$$

and

$$\|\boldsymbol{E}\| \le \|\boldsymbol{E}_1\| + \|\boldsymbol{E}_2\| \le \frac{15}{16}\epsilon + \frac{1}{16}\epsilon = \epsilon.$$

The proof is finished by invoking Lemma A.2 and utilizing the fact that $\boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*$ has only eigenvalues 1 and 0.

A.8 Proof of Corollary 3.5

We first consider the spectrum of $[\boldsymbol{F}_M]_L$, the top-left principal submatrix of $\boldsymbol{F}_M$. It is clear that its singular values are between 0 and 1 since it is a submatrix of $\boldsymbol{F}_M$. We first

observe that the gram matrix of $[\boldsymbol{F}_M]_L$, $[\boldsymbol{F}_M]_L^*[\boldsymbol{F}_M]_L$ is identical to $[\overline{\boldsymbol{B}}_{M,1/2p}]_L$ up to unitary phase factors, i.e.,

$$\begin{aligned}
&\left([\boldsymbol{F}_M]_L^*[\boldsymbol{F}_M]\right)[m,n] = e^{-j\pi\frac{m-n}{L-1}}\frac{\sin\left(\pi(m-n)/p\right)}{M\sin\left((m-n)\pi/M\right)} \\
&= e^{-j\pi\frac{m-n}{L-1}}\left([\overline{\boldsymbol{B}}_{M,1/2p}]_L\right)[m,n],\ \forall\ m,n\in[L].
\end{aligned}$$

This implies $[\boldsymbol{F}_M]_L^*[\boldsymbol{F}_M]_L$ has the same eigenvalue distribution to $[\overline{\boldsymbol{B}}_{M,1/2p}]_L$. Thus, Corollary 3.5 holds for $[\boldsymbol{F}_M]_L$ trivially by following Theorem 3.2.

Now note that any submatrix $\boldsymbol{F}_{M|p}$ obtained by deleting any consecutive $M-L$ columns and any consecutive $M-L$ rows of $\boldsymbol{F}_M$ is identical to $[\boldsymbol{F}_M]_L$ up to unitary phase factors

$$\boldsymbol{F}_{M|p} = \operatorname{diag}\left(\boldsymbol{a}_\xi\right)[\boldsymbol{F}_M]_L\operatorname{diag}\left(\boldsymbol{a}_\eta\right),$$

where ξ,η depend on the locations of the submatrix $\boldsymbol{F}_{M|p}$ and

$$\boldsymbol{a}_\xi = \begin{bmatrix}1 & e^{-2\pi\frac{\xi}{M}} & \cdots & e^{-2\pi\frac{(L-1)\xi}{M}}\end{bmatrix}^{\mathrm{T}}.$$

Thus, any submatrix $\boldsymbol{F}_{M|p}$ has the same spectrum as $[\boldsymbol{F}_M]_L$.

APPENDIX B

PROOFS FOR CHAPTER 4

This appendix mainly includes the proofs for Chapter 4.

B.1 Supporting Results

We establish several results which will also be useful in the remaining proofs. We start by the following result, a variant of Von Neumann's trace inquality [96].

Lemma B.1. *[96] For any $M \times N$ (suppose $M \leq N$) matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with singular values $\alpha_0 \geq \alpha_1 \geq \cdots \geq \alpha_{M-1} \geq 0$ and $\beta_0 \geq \beta_1 \geq \cdots \geq \beta_{M-1} \geq 0$, we have*

$$|\mathrm{trace}(\boldsymbol{A}\boldsymbol{B}^*)| \leq \sum_{m=0}^{M-1} \alpha_m \beta_m.$$

Proof of Lemma B.1. We enlarge $\boldsymbol{A}$ and $\boldsymbol{B}$ into $N \times N$ matrices $\boldsymbol{A}'$ and $\boldsymbol{B}'$ with zero rows, i.e.,

$$\boldsymbol{A}' = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{B}' = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0} \end{bmatrix}.$$

Let $\alpha'_0 \geq \alpha'_1 \geq \cdots \alpha'_{N-1}$ and $\beta'_0 \geq \beta'_1 \geq \cdots \geq \beta'_{N-1}$ be the singular values of $\boldsymbol{A}'$ and $\boldsymbol{B}'$, respectively. Note that $\alpha_n = \alpha'_n, \beta_n = \beta'_n$ for all $n \leq M-1$ and $\alpha'_n = 0, \beta'_n = 0$ for all $n \geq M$. It follows from Von Neumann's trace inquality [96] that

$$|\mathrm{trace}(\boldsymbol{A}\boldsymbol{B}^*)| = |\mathrm{trace}(\boldsymbol{A}'(\boldsymbol{B}')^*)| \leq \sum_{n=0}^{N-1} \alpha'_n \beta'_n = \sum_{m=0}^{M-1} \alpha_m \beta_m.$$

□

The following result provides an upper bound on $\varrho(\boldsymbol{Q})$ by the singular values of $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}$.

Lemma B.2. *Let $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW \rfloor - 1)\times R}$ be an orthonormal basis with $R \leq (N - 2\lfloor NW \rfloor - 1)$. Let $\pi_0 \geq \pi_1 \geq \cdots \geq \pi_{N-2\lfloor NW \rfloor - 2}$ denote the singular values of $\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}^*_{N,W}\boldsymbol{B}_{N,W}$. Then*

$$\varrho(\boldsymbol{Q}) = \int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2 \, df \leq \sum_{l=0}^{N-2\lfloor NW \rfloor - 2} \pi_l,$$

where $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix}$.

Proof of Lemma B.2. Recall (4.3) that

$$\varrho(\boldsymbol{Q}) = \int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2 \, df = \operatorname{trace}\left(\boldsymbol{B}_{N,W} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{B}_{N,W}\right).$$

Plugging in $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix}$, we have

$$\begin{aligned}
\varrho(\boldsymbol{Q}) &= \operatorname{trace}\left(\boldsymbol{B}_{N,W} - \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix}^* \boldsymbol{B}_{N,W}\right) \\
&= \operatorname{trace}\left(\boldsymbol{F}_N^*\boldsymbol{B}_{N,W}\boldsymbol{F}_N - \boldsymbol{F}_N^* \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix}^* \boldsymbol{B}_{N,W}\boldsymbol{F}_N\right) \\
&= \operatorname{trace}\left(\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}\right) \\
&\leq \sum_{l=0}^{N-2\lfloor NW \rfloor - 2} \pi_l \left\|\overline{\boldsymbol{F}}_{N,W}\right\| \\
&\leq \sum_{l=0}^{N-2\lfloor NW \rfloor - 2} \pi_l,
\end{aligned} \tag{B.1}$$

where the first inequality follows from Lemma B.1 by setting $\boldsymbol{A} = \overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$ and $\boldsymbol{B} = \overline{\boldsymbol{F}}_{N,W}^*$. □

In order to utilize Lemma B.2, we need the distribution of the singular values of $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$. This is established by the following result, whose proof is given in Appendix B.6.

Lemma B.3. *(singular value decay) Let* $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{N-2\lfloor NW \rfloor - 2}$ *denote the singular values of* $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$. *Then*

$$\sigma_\ell \leq \epsilon$$

when $\ell = C_N \log\left(\frac{15}{\epsilon}\right)$ *for any* $\epsilon \in (0,1)$. *Also*

$$\sigma_\ell \leq 15e^{-\frac{\ell}{C_N}}.$$

Now we are well equipped to prove the main results.

B.2 Proof of Theorem 4.1

We first provide the following results on the representation guarantee for the leading DPSS vectors and the subspace angle between the column spaces of $\boldsymbol{S}_K$ and $\boldsymbol{Q}$. Its proof is given in Appendix B.7.

Lemma B.4. *Let $\boldsymbol{V} \in \mathbb{C}^{(N-2\lfloor NW \rfloor -1)\times R}$ be an orthonormal basis with $R \leq (N - 2\lfloor NW \rfloor - 1)$. For any $\epsilon \in (0, \frac{1}{2})$, fix K to be such that $\lambda_{N,W}^{(K-1)} \geq \epsilon$. Then the orthobasis $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V} \end{bmatrix}$ satisfies*

$$\begin{aligned}
&\|\boldsymbol{S}_K\boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\boldsymbol{S}_K^*\|^2 \leq \eta := \frac{\left\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\right\|}{\epsilon}, \\
&\cos\left(\Theta_{\boldsymbol{S}_K,\boldsymbol{Q}}\right) \geq \sqrt{1 - N\eta}, \\
&\left\|\boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{s}_{N,W}^{(\ell)}\right\|_2^2 \leq \eta
\end{aligned}$$

for all $l = 0, 1, \ldots, K-1$.

Since $\boldsymbol{V}$ is the first R-principal eigenvectors of $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$, using Lemma B.3, we obtain

$$\left\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\right\| \leq 15e^{-\frac{R}{C_N}}.$$

If we set $R = C_N \log\left(\frac{15}{\epsilon^2}\right)$, we have

$$\left\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\right\| \leq \epsilon^2.$$

Alternatively, if one set $R = C_N \log\left(\frac{15N}{\epsilon^2}\right)$:

$$\left\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\right\| \leq \frac{\epsilon^2}{N}.$$

The proof is completed by utilizing Lemma B.4.

B.3 Proof of Theorem 4.2

Let $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{N-2\lfloor NW \rfloor -2}$ denote the singular values of $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$. Since $\boldsymbol{V}$ consists of the R dominant left singular vector of $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$, the singular values of $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}$ are $\sigma_R, \sigma_{R+1}, \ldots, \sigma_{N-2\lfloor NW \rfloor -2}$ and R zeros. It follows from Lemma B.3 that

$$\begin{aligned}\sum_{\ell=R}^{N-2\lfloor NW\rfloor-2}\sigma_\ell \leq & \sum_{\ell=R}^{N-2\lfloor NW\rfloor-2} 15e^{-\frac{\ell}{C_N}} \\ =&15\frac{e^{-\frac{R}{C_N}}(1-e^{-\frac{N-2\lfloor NW\rfloor-R-1}{C_N}})}{1-e^{-\frac{1}{C_N}}} \\ \leq&15\frac{e^{-\frac{R}{C_N}}}{1-e^{-\frac{1}{C_N}}} = 15\frac{e^{-\frac{R-1}{C_N}}}{e^{\frac{1}{C_N}}-1} \\ \leq&15e^{-\frac{R-1}{C_N}}C_N,\end{aligned} \tag{B.2}$$

where the last line follows from the Taylor expansion of $e^{\frac{1}{C_N}}$ which results in

$$\frac{1}{-1+1+\frac{1}{C_N}+\left(\frac{1}{C_N}\right)^2+\cdots} \leq C_N.$$

If $C_N \log\left(\frac{15C_N}{N\epsilon}\right)+1 \leq 0$, which implies that

$$\sum_{\ell=0}^{N-2\lfloor NW\rfloor-2}\sigma_\ell \leq N\epsilon,$$

thus by setting $R=0$ and $\boldsymbol{Q}=\boldsymbol{F}_{N,W}$ we are guaranteed that

$$\int_{-W}^{W}\frac{\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}df \leq \frac{1}{N}\sum_{\ell=0}^{N-2\lfloor NW\rfloor-2}\sigma_\ell \leq \frac{1}{N}N\epsilon=\epsilon.$$

Otherwise, choosing $R = C_N\log\left(\frac{15C_N}{N\epsilon}\right)+1$, we have

$$\sum_{\ell=R}^{N-2\lfloor NW\rfloor-2}\sigma_\ell \leq N\epsilon.$$

Now applying Lemma B.2, we have

$$\int_{-W}^{W}\frac{\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}df \leq \frac{1}{N}\sum_{\ell=R}^{N-2\lfloor NW\rfloor-2}\sigma_\ell \leq \frac{1}{N}N\epsilon=\epsilon,$$

where we utilize the fact that each sinusoid has energy $\|\boldsymbol{e}_f\|_2^2=N$.

Remark B.1. By (B.1), we have

$$\varrho(\boldsymbol{Q})=\text{trace}\left(\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}-\boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}\right).$$

Directly solving

$$\underset{\boldsymbol{V}\in\mathbb{C}^{(N-2\lfloor NW\rfloor-1)\times R}}{\text{minimize}} \operatorname{trace}\left(\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} - \boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}\right),$$

we obtain an alternative optimal solution $\boldsymbol{V}'$ consisting of the first R principal eigenvectors of $\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}$. The orthobasis $\boldsymbol{Q}' = \begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}'\end{bmatrix}$ is optimal in terms of minimizing $\varrho(\boldsymbol{Q})$ and also for averagely representing all discrete-time sinusoids with a frequency $f \in [-W, W]$ in the least square sense. We can also have similar approximation guarantee for $\boldsymbol{V}'$. Note that

$$\boldsymbol{F}_{N}^{*}\left(\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^{*}\right)\boldsymbol{F}_{N} = \begin{bmatrix} \boldsymbol{F}_{N,W}^{*}\boldsymbol{B}_{N,W}\boldsymbol{F}_{N,W} - \mathbf{I} & \boldsymbol{F}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} \\ \overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} \end{bmatrix}.$$

By utilizing the result that

$$\boldsymbol{B}_{N,W} = \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^{*} + \boldsymbol{L} + \boldsymbol{E},$$

where

$$\operatorname{rank}(\boldsymbol{L}) \leq C_N \log\left(\frac{15}{\epsilon}\right) \qquad \text{and} \qquad \|\boldsymbol{E}\| \leq \epsilon,$$

we can rewrite $\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} = \boldsymbol{L}_2 + \boldsymbol{E}_2$, where

$$\boldsymbol{L}_2 := \overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{L}\overline{\boldsymbol{F}}_{N,W} \qquad \text{and} \qquad \boldsymbol{E}_2 := \overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{E}\overline{\boldsymbol{F}}_{N,W}.$$

Thus,

$$\operatorname{rank}(\boldsymbol{L}_2) \leq C_N \log\left(\frac{15}{\epsilon}\right) \qquad \text{and} \qquad \|\boldsymbol{E}_2\| \leq \epsilon.$$

It follows from Eckart-Young-Mirsky theorem [44] that

$$\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} - \boldsymbol{V}'(\boldsymbol{V}')^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}\| \leq \|\boldsymbol{E}_2\| \leq \epsilon.$$

Therefore, choosing $R = C_N \log\left(\frac{15C_N}{N\epsilon}\right) + 1$, with similar argument we also have

$$\int_{-W}^{W} \frac{\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} df \leq \frac{1}{N}\operatorname{trace}\left(\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W} - \boldsymbol{V}'(\boldsymbol{V}')^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\overline{\boldsymbol{F}}_{N,W}\right) \leq \epsilon.$$

We note that all the other results on $\boldsymbol{V}$ can also be applied to $\boldsymbol{V}'$ with similar or slightly different guarantee.

B.4 Proof of Theorem 4.3

By Theorem 4.2, we are guaranteed that the pure sinusoids have, on average, a small representation residual in the basis $\boldsymbol{Q}$. Intuitively, the representation error for each pure sinusoids is also guaranteed to be small. The following result provides an upper bound on the representation error for each pure sinusoid with the average representation error. Its proof is given in Appendix B.8.

Lemma B.5. *For any $q \in \{1, 2, \ldots, N\}$, suppose $\boldsymbol{U} \in \mathbb{C}^{N\times q}$ is an orthonormal basis such that $\boldsymbol{U}^*\boldsymbol{U} = \mathbf{I}$. Also suppose $W \geq \frac{1}{4\pi N}$. Then*

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} \leq \max\left(2\sqrt{\pi}\sqrt{\int_{-W}^{W}\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2\,df},\ \frac{1}{NW}\int_{-W}^{W}\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2\right).$$

Proof of Theorem 4.3. It follows from (B.2) that by choosing $R = C_N \log\left(\frac{15C_N}{\epsilon'}\right) + 1$, we have

$$\sum_{l=R}^{N-2\lfloor NW\rfloor-1} \sigma_l \leq \epsilon'.$$

Utilizing Lemma B.2 gives

$$\int_{-W}^{W}\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2\,df \leq \sum_{l=R}^{N-2\lfloor NW\rfloor-1}\sigma_l \leq \epsilon'.$$

The proof is completed by setting

$$\epsilon' = \frac{\epsilon^2}{4\pi}, \quad R = C_N \log\left(\frac{60\pi C_N}{\epsilon^2}\right) + 1,$$

or

$$\epsilon' = NW\epsilon, \quad R = C_N \log\left(\frac{15C_N}{NW\epsilon}\right) + 1.$$

□

B.5 Proof of Theorem 4.5

We first present the following guarantees on randomized algorithms for computing orthonormal basis from [60].

Theorem B.1. *[60, Theorem 10.5] (average Frobenius norm) Let $\boldsymbol{A}$ be an $M \times N$ (suppose $M \leq N$) matrix with singular values $\alpha_0 \geq \alpha_1 \geq \cdots \geq \alpha_{M-1}$. Choose a target rank $R \geq 2$ and an oversampling parameter $p \geq 2$, where $P = R + p \leq M$. Let $\boldsymbol{\Omega}$ be an $N \times P$ standard Gaussian matrix. Let $\boldsymbol{P_Y}$ be an orthogonal projector onto the column space of the sample matrix $\boldsymbol{Y} = \boldsymbol{A\Omega}$. Then the expected approximation error*

$$\mathbb{E}\left[\|\boldsymbol{A} - \boldsymbol{P_Y A}\|_F\right] \leq \left(1 + \frac{R}{p-1}\right)^{1/2} \left(\sum_{m=R}^{M-1} \alpha_m^2\right)^{1/2},$$

where $\mathbb{E}$ denotes expectation with respect to the random matrix $\boldsymbol{\Omega}$.

Theorem B.2. *[60, Theorem10.6] (average spectral error) Under the setup of Theorem B.1,*

$$\mathbb{E}\left[\|\boldsymbol{A} - \boldsymbol{P_Y A}\|\right] \leq \left(1 + \sqrt{\frac{R}{p-1}}\right) \alpha_R + \frac{e\sqrt{P}}{p} \left(\sum_{m=R}^{M-1} \alpha_m^2\right)^{1/2}.$$

Proof of Theorem 4.5. Let $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{N-2\lfloor NW \rfloor - 2}$ denote the singular values of $\overline{\boldsymbol{F}}^*_{N,W} \boldsymbol{B}_{N,W}$. Utilizing Lemma B.3, we have

$$\begin{aligned}
\sum_{l=R}^{N-2\lfloor NW \rfloor - 2} \sigma_l^2 \leq & \sum_{l=R}^{N-2\lfloor NW \rfloor - 2} \left(15 e^{-\frac{\ell}{C_N}}\right)^2 \\
= & 225 \frac{e^{-2\frac{R}{C_N}} \left(1 - e^{-2\frac{N-2\lfloor NW \rfloor - R - 1}{C_N}}\right)}{1 - e^{-\frac{2}{C_N}}} \\
\leq & 225 \frac{e^{-2\frac{R}{C_N}}}{1 - e^{-\frac{2}{C_N}}} = 225 \frac{e^{-2\frac{R-1}{C_N}}}{e^{\frac{2}{C_N}} - 1} \\
\leq & 225 e^{-2\frac{R-1}{C_N}} \frac{C_N}{2}.
\end{aligned}$$

Note that here $\boldsymbol{V}$ is an orthonormal basis for the column space of the sample matrix $\overline{\boldsymbol{F}}^*_{N,W} \boldsymbol{B}_{N,W} \boldsymbol{\Omega}$. Let $\pi_0 \geq \pi_1 \geq \cdots \geq \pi_{N-2\lfloor NW \rfloor - 2}$ denote the singular values of $\overline{\boldsymbol{F}}^*_{N,W} \boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^* \overline{\boldsymbol{F}}^*_{N,W} \boldsymbol{B}_{N,W}$.

- Utilizing Theorem B.2, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|\right]\\
&\leq\left(1+\sqrt{\frac{R}{P-R-1}}\right)\sigma_R+\frac{e\sqrt{P}}{P-R}\left(\sum_{l=R}^{M-1}\sigma_l^2\right)^{1/2}\\
&\leq\left(1+\sqrt{\frac{R}{P-R-1}}\right)15e^{-\frac{R}{C_N}}+\frac{e\sqrt{P}}{P-R}\left(225e^{-2\frac{R-1}{C_N}}\frac{C_N}{2}\right)^{1/2}\\
&=\left(1+\sqrt{\frac{R}{P-R-1}}\right)15e^{-\frac{R}{C_N}}+15\frac{e\sqrt{P}}{P-R}e^{-\frac{R-1}{C_N}}\sqrt{\frac{C_N}{2}}.
\end{aligned}$$

  Setting $R=C_N\log\left(\frac{30+15e}{\epsilon^2}\right)+1$ and $P=2R+1$, we have

$$\mathbb{E}\left[\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|\right]\leq 30\frac{\epsilon^2}{30+15e}+15e\sqrt{\frac{C_N}{R+1}}\frac{\epsilon^2}{30+15e}\leq\epsilon^2$$

  since $C_N\leq R$ for any $\epsilon^2\in(0,1)$. It follows from Lemma B.4 that

$$\begin{aligned}
&\mathbb{E}\left[\|\boldsymbol{S}_K\boldsymbol{S}_K^{*}-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{S}_K\boldsymbol{S}_K^{*}\|^2\right]\leq\frac{\mathbb{E}\left[\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|\right]}{\epsilon}\leq\epsilon\\
&\mathbb{E}\left[\|\boldsymbol{s}_{N,W}^{(\ell)}-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{s}_{N,W}^{(\ell)}\|^2\right]\leq\epsilon
\end{aligned}$$

  for all $l=0,1,\ldots,K-1$. Alternatively, setting $R=C_N\log\left(\frac{(30+15e)N}{\epsilon^2}\right)+1$ and $P=2R+1$, we have

$$\mathbb{E}\left[\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|\right]\leq\frac{\epsilon^2}{N}.$$

  Thus applying Lemma B.4 gives

$$\mathbb{E}\left[\cos\left(\Theta_{\boldsymbol{S}_K,\boldsymbol{Q}}\right)\right]\geq\sqrt{1-N\frac{\mathbb{E}\left[\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|\right]}{\epsilon}}\geq\sqrt{1-\epsilon}.$$

- Set $p=\frac{R}{3}+1$, i.e., $P=\frac{4}{3}R+1$. It follows from Theorem B.1 that

$$\begin{aligned}\mathbb{E}\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|_{F}&\leq\left(1+\frac{R}{p-1}\right)^{1/2}\left(\sum_{l=R}^{N-2\lfloor NW\rfloor-2}\sigma_{l}^{2}\right)^{1/2}\\&\leq 2\sqrt{225e^{-2\frac{R-1}{C_N}}\frac{C_N}{2}}\\&=15e^{-\frac{R-1}{C_N}}\sqrt{2C_N}.\end{aligned}$$

By applying Lemma B.2 and utilizing the inequality between the Frobenius norm and the nuclear norm, we have

$$\begin{aligned}&\mathbb{E}\int_{-W}^{W}\frac{\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}df=\frac{1}{N}\mathbb{E}\sum_{m=0}^{N-2\lfloor NW\rfloor-2}\pi_m\\&\leq\frac{1}{N}N\mathbb{E}\left\|\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}-\boldsymbol{V}\boldsymbol{V}^{*}\overline{\boldsymbol{F}}_{N,W}^{*}\boldsymbol{B}_{N,W}\right\|_{F}\leq 15e^{-\frac{R-1}{C_N}}\sqrt{2C_N}.\end{aligned}\tag{B.3}$$

Setting $R=C_N\log\left(\frac{15\sqrt{2C_N}}{\epsilon}\right)+1$, we obtain

$$\mathbb{E}\int_{-W}^{W}\frac{\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}df\leq\epsilon.$$

- Set $p=\frac{R}{3}+1$, i.e., $P=\frac{4}{3}R+1$. From (B.3), it follows that

$$\mathbb{E}\left[\int_{-W}^{W}\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2\right]\leq 15Ne^{-\frac{R-1}{C_N}}\sqrt{2C_N}.$$

Utilizing Lemma B.5, we have

$$\begin{aligned}&\mathbb{E}\left[\frac{\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}\right]\\&\leq\max\left(\mathbb{E}\left[2\sqrt{\pi}\sqrt{\int_{-W}^{W}\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2\,df}\right],\mathbb{E}\left[NW\int_{-W}^{W}\|\boldsymbol{e}_f-\boldsymbol{Q}\boldsymbol{Q}^{*}\boldsymbol{e}_f\|_2^2\right]\right)\\&\leq\max\left(2\sqrt{15\pi N\sqrt{2C_N}}e^{-\frac{R-1}{2C_N}},\frac{15Ne^{-\frac{R-1}{C_N}}\sqrt{2C_N}}{NW}\right).\end{aligned}$$

Setting

$$R = \max\left(C_N \log\left(\frac{60\pi N\sqrt{2C_N}}{\epsilon^2}\right) + 1, C_N \log\left(\frac{15\pi\sqrt{2C_N}}{W\epsilon}\right) + 1\right)$$

yields

$$\mathbb{E}\left[\frac{\|\boldsymbol{e}_f - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2}\right] \leq \epsilon.$$

□

B.6 Proof of Lemma B.3

Note that

$$\boldsymbol{F}_N^*\left(\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^*\right) = \begin{bmatrix} \boldsymbol{F}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{F}_{N,W}^* \\ \overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} \end{bmatrix}.$$

By utilizing the result that

$$\boldsymbol{B}_{N,W} = \boldsymbol{F}_{N,W}\boldsymbol{F}_{N,W}^* + \boldsymbol{L} + \boldsymbol{E},$$

where

$$\operatorname{rank}(\boldsymbol{L}) \leq C_N \log\left(\frac{15}{\epsilon}\right) \qquad \text{and} \qquad \|\boldsymbol{E}\| \leq \epsilon,$$

we can rewrite $\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} = \boldsymbol{L}_1 + \boldsymbol{E}_1$, where

$$\boldsymbol{L}_1 := \overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{L} \qquad \text{and} \qquad \boldsymbol{E}_1 := \overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{E}.$$

Thus,

$$\operatorname{rank}(\boldsymbol{L}_1) \leq C_N \log\left(\frac{15}{\epsilon}\right) \qquad \text{and} \qquad \|\boldsymbol{E}_1\| \leq \epsilon.$$

It follows from Eckart-Young-Mirsky theorem [44] that

$$\sigma_{\operatorname{rank}(\boldsymbol{L}_1)} \leq \|\boldsymbol{E}_1\| \leq \epsilon$$

for any $\epsilon \in (0, 1)$. Noting that $\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\| \leq \|\overline{\boldsymbol{F}}_{N,W}^*\|\|\boldsymbol{B}_{N,W}\| < 1$, we have

$$\sigma_\ell \leq 15e^{-\frac{\ell}{C_N}}.$$

for all $\ell = 0, 1, \ldots, N - 2\lfloor NW \rfloor - 2$. Otherwise, suppose $\sigma_\ell > 15e^{-\frac{\ell}{C_N}}$. If $15e^{-\frac{\ell}{C_N}} \geq 1$, then it contradicts to the fact that $\sigma_\ell < 1$. If $15e^{-\frac{\ell}{C_N}} < 1$, let $\epsilon = 15e^{-\frac{\ell}{C_N}}$. Then we have a contradiction to the fact that $\sigma_{\text{rank}(\boldsymbol{L}_1)} \leq \epsilon$ and $\text{rank}(\boldsymbol{L}_1) \leq C_N \log\left(\frac{15}{\epsilon}\right) = \ell$.

B.7 Proof of Lemma B.4

Fix K to be such that $\lambda_{N,W}^{(K-1)} > \epsilon$. Utilizing $\boldsymbol{B}_{N,W} = \boldsymbol{S}_{N,W}\boldsymbol{\Lambda}_{N,W}\boldsymbol{S}_{N,W}^*$, we have

$$\begin{aligned}
\|\boldsymbol{B}_{N,W} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{B}_{N,W}\| =& \|\boldsymbol{S}_{N,W}\boldsymbol{\Lambda}_{N,W}\boldsymbol{S}_{N,W}^* - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_{N,W}\boldsymbol{\Lambda}_{N,W}\boldsymbol{S}_{N,W}^*\| \\
=& \|\boldsymbol{\Lambda}_{N,W} - \boldsymbol{S}_{N,W}^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_{N,W}\boldsymbol{\Lambda}_{N,W}\| \\
\geq& \|\boldsymbol{\Lambda}_K - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*K\boldsymbol{\Lambda}_K\| \\
=& \|\left(\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\right)\boldsymbol{\Lambda}_K\| \\
\geq& \|\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\|\,\epsilon.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\|\boldsymbol{B}_{N,W} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{B}_{N,W}\| &= \left\|\boldsymbol{B}_{N,W} - \begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}^*\boldsymbol{B}_{N,W}\right\| \\
&= \left\|\boldsymbol{F}_N^*\boldsymbol{B}_{N,W} - \boldsymbol{F}_N^*\begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{N,W} & \overline{\boldsymbol{F}}_{N,W}\boldsymbol{V}\end{bmatrix}^*\boldsymbol{B}_{N,W}\right\| \\
&= \left\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\right\|.
\end{aligned}$$

Combining the above two sets of equations yields

$$\|\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\| \leq \eta = \frac{\left\|\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W} - \boldsymbol{V}\boldsymbol{V}^*\overline{\boldsymbol{F}}_{N,W}^*\boldsymbol{B}_{N,W}\right\|}{\epsilon}.$$

Now exploit the relationship between $\boldsymbol{S}_K\boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\boldsymbol{S}_K^*$ and $\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K$ as follows

$$\begin{aligned}
\|\boldsymbol{S}_K\boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\boldsymbol{S}_K^*\|^2 &= \left\|\left(\boldsymbol{S}_K\boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\boldsymbol{S}_K^*\right)^{\mathrm{T}}\left(\boldsymbol{S}_K\boldsymbol{S}_K^* - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\boldsymbol{S}_K^*\right)\right\| \\
&= \|\boldsymbol{S}_K\left(\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\right)\boldsymbol{S}_K^\star\| \\
&\leq \|(\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K)\| \leq \eta.
\end{aligned}$$

Then, utilizing the inequality $\|\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\|_{\max} \leq \|\mathbf{I} - \boldsymbol{S}_K^*\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{S}_K\|$, where $\|\boldsymbol{A}\|_{\max}$ is the maximum absolute entries of $\boldsymbol{A}$, we have

$$\left|\left(\boldsymbol{s}_{N,W}^{(\ell)}\right)^* \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}_{N,W}^{(l')}\right| \leq \|\mathbf{I} - \boldsymbol{S}_K^* \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{S}_K\| \leq \eta$$

for all $\ell \neq \ell', \ell, \ell' = 0, 1, \ldots, K-1$, and

$$\left\|\boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}_{N,W}^{(\ell)}\right\|_2^2 = 1 - \left\|\boldsymbol{Q}^* \boldsymbol{s}_{N,W}^{(\ell)}\right\|_2^2 \leq \|\mathbf{I} - \boldsymbol{S}_K^* \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{S}_K\| \leq \eta$$

for all $\ell = 0, 1, \ldots, K-1$.

Let $\boldsymbol{s}$ be an arbitrary unit vector in the subspace spanned by $\boldsymbol{S}_K$, i.e., $\boldsymbol{s} = \sum_{\ell=0}^{K-1} \alpha_\ell \boldsymbol{s}_{N,W}^{(\ell)}$ with $\|\boldsymbol{s}\|_2 = \sum_{\ell=0}^{K-1} \alpha_\ell^2 = 1$. We have

$$\begin{aligned}
\|\boldsymbol{s} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}\|_2 &= \left\|\sum_{\ell=0}^{K-1} \alpha_\ell \left(\boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}_{N,W}^{(\ell)}\right)\right\|_2 \\
&\leq \sum_{\ell=0}^{K-1} |\alpha_\ell| \left\|\boldsymbol{s}_{N,W}^{(\ell)} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}_{N,W}^{(\ell)}\right\|_2 \\
&\leq \sqrt{\eta} \sum_{\ell=0}^{K-1} |\alpha_\ell| \\
&\leq \sqrt{K\eta} \leq \sqrt{N\eta}
\end{aligned}$$

where the last line follows from the inequality between L^1-norm and L^2-norm: $\|\boldsymbol{a}\|_1 \leq \sqrt{K}\|\boldsymbol{a}\|_2$ for any $\boldsymbol{a} \in \mathbb{R}^K$. Thus, we obtain

$$\|\boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}\|_2^2 = 1 - \|\boldsymbol{s} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{s}\|_2^2 \geq 1 - N\eta.$$

Since this result holds for arbitrary unit vector $\boldsymbol{s}$ in the subspace spanned by $\boldsymbol{S}_K$, we finally have

$$\cos(\Theta_{\boldsymbol{S}_K, \boldsymbol{Q}}) \geq \sqrt{1 - N\eta}.$$

B.8 Proof of Lemma B.5

Let $\boldsymbol{\Phi}$ be an $N \times N$ diagonal matrix with diagonal entries $j2\pi 0, j2\pi, \ldots, j2\pi(N-1)$. The derivative of $\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{e}_f\|_2^2$ in terms of f can be computed as

$$\frac{d}{df}\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2=2\mathbb{R}\left(\boldsymbol{e}_f^*\left(\mathbf{I}-\boldsymbol{U}\boldsymbol{U}^*\right)\boldsymbol{\Pi}\boldsymbol{e}_f\right).$$

We first obtain an upper bound for its derivative

$$\begin{aligned}\left|\frac{d}{df}\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2\right| \leq & 2\left|\boldsymbol{e}_f^*\left(\mathbf{I}-\boldsymbol{U}\boldsymbol{U}^*\right)\boldsymbol{\Pi}\boldsymbol{e}_f\right| \\ \leq & 2\left|\boldsymbol{e}_f^*\boldsymbol{\Pi}\boldsymbol{e}_f\right|\left\|\mathbf{I}-\boldsymbol{U}\boldsymbol{U}^*\right\| \\ \leq & 2\left|\boldsymbol{e}_f^*\boldsymbol{\Pi}\boldsymbol{e}_f\right| \leq 2\pi N(N-1) \leq 2\pi N^2\end{aligned}$$

for all $f \in [0,1]$. Since $\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2$ is nonngetaive and its derivative is bounded above, $\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2$ cannot be too large if $\int_{-W}^{W}\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2 df$ is very small.

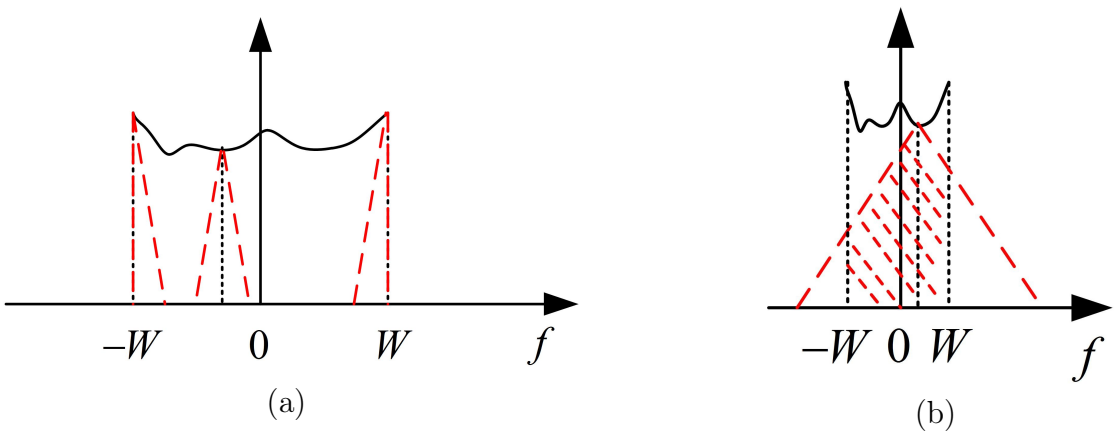


Figure B.1: (a) Illustration of (B.4), where the area below the black curve is always larger than or equal to the area of each red triangle; (b) Illustration of (B.5), where the area below the black curve is always larger than or equal to the area indicated by red dashed lines.

Suppose $\frac{\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2}{2\pi N^2} \leq 2W$. As illustrated in Figure B.1, for any $f \in [-W,W]$, we can always find a triangle with area either

$$\frac{\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^4}{2\sup_{f\in[-W,W]}\left|\frac{d}{df}\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2\right|}$$

(the area of the left and right red triangles) or

$$\frac{\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^4}{\sup_{f\in[-W,W]}\left|\frac{d}{df}\left\|\boldsymbol{e}_f-\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\right\|_2^2\right|}$$

(the area of the middle red triangle) that is smaller than $\int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \, df$ (the area under the black curve). This is made more precise as

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^4}{4\pi N^2} \leq \frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^4}{2\sup_{f\in[-W,W]}\left|\frac{d}{df}\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2\right|} \leq \int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \, df \quad \text{(B.4)}$$

for all $f \in [-W, W]$. Thus, we have

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} = \frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2}{N} \leq 2\sqrt{\pi}\sqrt{\int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \, df}$$

for all $f \in [-W, W]$.

On the other hand, suppose $\frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2}{2\pi N^2} > 2W$. With similar argument, as illustrated in Figure B.1, for any $f \in [-W, W]$, we can always find a region of area at least $W\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2$ (the area indicated by red dashed lines) that is smaller than that is smaller than $\int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \, df$ (the area under the black curve). This is made more precise as

$$W\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \leq \int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \, df \quad \text{(B.5)}$$

for all $f \in [-W, W]$. Thus, we have

$$\frac{\|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2}{\|\boldsymbol{e}_f\|_2^2} \leq \frac{1}{NW}\int_{-W}^{W} \|\boldsymbol{e}_f - \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e}_f\|_2^2 \, df$$

for all $f \in [-W, W]$.

APPENDIX C

PROOFS FOR CHAPTER 5

This appendix contains the proofs for Chapter 5.

C.1 Proof of Lemma 5.1

Let $\boldsymbol{y} \in \mathbb{C}^N, \boldsymbol{y} \neq \boldsymbol{0}$ be an arbitrary vector. Then

$$\langle \mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y}))), \boldsymbol{y} \rangle = \sum_{m=0}^{N-1} \mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y})))[m]\overline{\boldsymbol{y}}[m] = \sum_{m=0}^{N-1} \left(\sum_{n=0}^{N-1} \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \boldsymbol{y}[n] \right) \overline{\boldsymbol{y}}[m]$$
$$= \int_{\mathbb{W}} \left(\sum_{m=0}^{N-1} e^{j2\pi fm} \overline{\boldsymbol{y}}[m] \right) \left(\sum_{n=0}^{N-1} e^{-j2\pi fn} \boldsymbol{y}[n] \right) df = \int_{\mathbb{W}} |\sum_{n=0}^{N-1} \boldsymbol{y}[n] e^{-j2\pi fn}|^2 df > 0,$$

where $\overline{\boldsymbol{y}}$ is the complex-conjugate of the vector $\boldsymbol{y}$, $\sum_{n=0}^{N-1} \boldsymbol{y}[n] e^{-j2\pi fn}$ is the DTFT of $\mathcal{I}_N^*(\boldsymbol{y})$, and the last inequality is derived from the fact that compactly supported signals cannot have perfectly flat magnitude response.

By Parsevel's Theorem, we know $\int_{-1/2}^{1/2} |\sum_{n=0}^{N-1} \boldsymbol{y}[n] e^{-j2\pi fn}|^2 df = ||\boldsymbol{y}||_2^2$. Therefore

$$\langle \mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y}))), \boldsymbol{y} \rangle = \int_{\mathbb{W}} |\sum_{n=0}^{N-1} \boldsymbol{y}[n] e^{-j2\pi fn}|^2 df < ||\boldsymbol{y}||_2^2.$$

Thus, we have

$$0 < \min_{\boldsymbol{y} \in \mathbb{C}^N} \frac{\langle \mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y}))), \boldsymbol{y} \rangle}{||\boldsymbol{y}||_2^2} \leq \lambda_{N,\mathbb{W}}^{(\ell)} \leq \max_{\boldsymbol{y} \in \mathbb{C}^N} \frac{\langle \mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{y}))), \boldsymbol{y} \rangle}{||\boldsymbol{y}||_2^2} < 1$$

for all $\ell \in [N]$.

By noting that $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ is equivalent to $\boldsymbol{B}_{N,\mathbb{W}}$, we have

$$\sum_{\ell=0}^{N-1} \lambda_{N,\mathbb{W}}^{(\ell)} = \text{trace}(\boldsymbol{B}_{N,\mathbb{W}}) = \sum_{n=0}^{N-1} \boldsymbol{B}_{N,\mathbb{W}}[n,n] = \sum_{n=0}^{N-1} \int_{\mathbb{W}} e^{j2\pi f 0} df = N|\mathbb{W}|.$$

C.2 Proof of Theorem 5.1

First we state a useful inequality about the Frobenius norm of positive semi-definite matrices. Suppose $\boldsymbol{X} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{Y} \in \mathbb{C}^{N\times N}$ are two arbitrary positive semi-definite matrices. Then

$$\begin{aligned}||\boldsymbol{X}+\boldsymbol{Y}||_F^2 &= \text{trace}\big((\boldsymbol{X}+\boldsymbol{Y})^H(\boldsymbol{X}+\boldsymbol{Y})\big)\\ &= ||\boldsymbol{X}||_F^2 + ||\boldsymbol{Y}||_F^2 + 2\text{trace}(\boldsymbol{X}^H\boldsymbol{Y})\\ &\geq ||\boldsymbol{X}||_F^2 + ||\boldsymbol{Y}||_F^2,\end{aligned}$$

where the last inequality is derived from the fact that $\text{trace}(\boldsymbol{X}^H\boldsymbol{Y})$ is nonnegative, which can be showed as follows. By the hypothesis that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are positive semi-definite matrices, we have the factorization $\boldsymbol{X}^H = \boldsymbol{X} = \boldsymbol{X}^{1/2}\boldsymbol{X}^{1/2}$, where $\boldsymbol{X}^{1/2}$ is also a positive semi-definite matrix.[20] Then we conclude that $\text{trace}(\boldsymbol{X}^H\boldsymbol{Y}) = \text{trace}(\boldsymbol{X}^{1/2}\boldsymbol{X}^{1/2}\boldsymbol{Y}) = \text{trace}(\boldsymbol{X}^{1/2}\boldsymbol{Y}\boldsymbol{X}^{1/2}) \geq 0$, since $\boldsymbol{X}^{1/2}\boldsymbol{Y}\boldsymbol{X}^{1/2}$ is also a positive semi-definite matrix.

We next bound the Frobenius norm of $\boldsymbol{B}_{N,W_i}$ by

$$\begin{aligned}||\boldsymbol{B}_{N,W_i}||_F^2 &= N(2W_i)^2 + \sum\sum_{m\neq n}\left(\frac{\sin\left(2\pi W_i(m-n)\right)}{\pi(m-n)}\right)^2\\ &= 4NW_i^2 + 2\sum_{n=1}^{N-1}(N-n)\left(\frac{\sin\left(2\pi W_i n\right)}{\pi n}\right)^2\\ &= 4NW_i^2 + 2N\sum_{n=1}^{N-1}\left(\frac{\sin\left(2\pi W_i n\right)}{\pi n}\right)^2 - 2\sum_{n=1}^{N-1} n\left(\frac{\sin\left(2\pi W_i n\right)}{\pi n}\right)^2\\ &= 4NW_i^2 + 2N\left(W_i - 2W_i^2 - \sum_{n=N}^{\infty}\left(\frac{\sin\left(2\pi W_i n\right)}{\pi n}\right)^2\right) - 2\sum_{n=1}^{N-1} n\left(\frac{\sin\left(2\pi W_i n\right)}{\pi n}\right)^2\\ &\geq 4NW_i^2 + 2N\left(W_i - 2W_i^2 - \frac{1}{\pi^2}\int_{N-1}^{\infty}\frac{1}{x^2}dx\right) - 2\frac{1}{\pi^2}\left(\int_1^{N-1}\frac{1}{x}dx + 1\right)\\ &= 2NW_i - \frac{2}{\pi^2}\frac{2N-1}{N-1} - \frac{2}{\pi^2}\log(N-1),\end{aligned}$$

[20] Note that $\boldsymbol{X}$ has the eigen-decomposition $\boldsymbol{X} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H$ where $\boldsymbol{V}$ is an orthonormal matrix and $\boldsymbol{D}$ is a diagonal matrix whose diagonal elements are non-negative, giving the square root $\boldsymbol{X}^{1/2} = \boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{V}^H$.

where the fourth line follows from Parseval's theorem $\sum_{n=-\infty}^{\infty}\left(\frac{\sin(2\pi W_i n)}{\pi n}\right)^2 = \int_{-W_i}^{W_i} df = 2W_i$, which indicates that $\sum_{n=1}^{\infty}\left(\frac{\sin(2\pi W_i n)}{\pi n}\right)^2 = W_i - 2W_i^2$.

Now applying the above results yields

$$\begin{aligned}
||\boldsymbol{B}_{N,\mathbb{W}}||_F^2 &= ||\sum_{i=0}^{J-1} \boldsymbol{E}_{f_i}\boldsymbol{B}_{N,W_i}\boldsymbol{E}_{f_i}^H||_F^2 \\
&\geq \sum_{i=0}^{J-1} ||\boldsymbol{B}_{N,W_i}||_F^2 \\
&\geq \sum_{i=0}^{J-1}\left(2NW_i - \frac{2}{\pi^2}\frac{2N-1}{N-1} - \frac{2}{\pi^2}\log(N-1)\right) \\
&= N|\mathbb{W}| - J\left(\frac{2}{\pi^2}\frac{2N-1}{N-1} + \frac{2}{\pi^2}\log(N-1)\right),
\end{aligned}$$

where the second line follows since $\boldsymbol{E}_{f_i}\boldsymbol{B}_{N,W_i}\boldsymbol{E}_{f_i}^H$ is positive semi-definite. Recalling the result stated in Lemma 5.1 that $\sum_{\ell=0}^{N-1}\lambda_{N,\mathbb{W}}^{(\ell)} = \text{trace}(\boldsymbol{B}_{N,\mathbb{W}}) = N|\mathbb{W}|$, we get

$$\sum_{\ell=0}^{N-1}\lambda_{N,\mathbb{W}}^{(\ell)}(1-\lambda_{N,\mathbb{W}}^{(\ell)}) = \text{trace}(\boldsymbol{B}_{N,\mathbb{W}}) - ||\boldsymbol{B}_{N,\mathbb{W}}||_F^2 \leq J\left(\frac{2}{\pi^2}\frac{2N-1}{N-1} + \frac{2}{\pi^2}\log(N-1)\right).$$

Thus, equation (5.2) follows by noting that for any $\varepsilon \in (0, \frac{1}{2})$ one has

$$\sum_{\ell=0}^{N-1}\lambda_{N,\mathbb{W}}^{(\ell)}(1-\lambda_{N,\mathbb{W}}^{(\ell)}) \geq \sum_{\{l:\varepsilon\leq\lambda_{N,\mathbb{W}}^{(\ell)}\leq 1-\varepsilon\}}\lambda_{N,\mathbb{W}}^{(\ell)}(1-\lambda_{N,\mathbb{W}}^{(\ell)}) \geq \varepsilon(1-\varepsilon)\#\{l : \varepsilon \leq \lambda_{N,\mathbb{W}}^{(\ell)} \leq 1-\varepsilon\}.$$

C.3 Proof of Theorem 5.2

A precise proof of a similar result for time- and band-limiting operators in the continuous domain was first given in [80]. Izu and Lakey [70] extend the result to multiple intervals in the frequency domain or time domain. Their work forms the foundation of the following analysis.

As we have noted, the two operators $\mathcal{T}_N\mathcal{B}_{\mathbb{W}}\mathcal{T}_N$ and $\mathcal{I}_N\mathcal{B}_{\mathbb{W}}\mathcal{I}_N^*$ have the same eigenvalues. We work with $\mathcal{T}_N\mathcal{B}_{\mathbb{W}}\mathcal{T}_N$ to prove Theorem 5.2. For convenience, we also use $\lambda_{N,\mathbb{W}}^{(0)}, \ldots, \lambda_{N,\mathbb{W}}^{(N-1)}$ to denote the decreasing eigenvalues for the operator $\mathcal{T}_N\mathcal{B}_{\mathbb{W}}\mathcal{T}_N$. We let $S([N])$ denote the

subspace of all finite-energy sequences supported only on the index set $[N]$, that is

$$S([N]) = \{y : y \in \ell_2(\mathbb{Z}), \mathcal{T}_N(y) = y\}.$$

First, for all integers $\ell \in [N]$, the Weyl-Courant minimax representation of the eigenvalues can be stated as

$$\begin{aligned}
\lambda_{N,\mathbb{W}}^{(\ell)} &= \begin{cases} \min_{S_\ell} \max_{y \in \ell_2(\mathbb{Z}), y \perp S_\ell} \frac{\langle \mathcal{T}_N(\mathcal{B}_\mathbb{W}(\mathcal{T}_N(y))), y\rangle}{\langle y, y\rangle}, \\ \max_{S_{l+1}} \min_{y \in \ell_2(\mathbb{Z}), y \in S_{l+1}} \frac{\langle \mathcal{T}_N(\mathcal{B}_\mathbb{W}(\mathcal{T}_N(y))), y\rangle}{\langle y, y\rangle}, \end{cases} \\
&= \begin{cases} \min_{S_\ell} \max_{y \in S([N]), y \perp S_\ell} \frac{\langle \mathcal{T}_N(\mathcal{B}_\mathbb{W}(\mathcal{T}_N(y))), y\rangle}{\langle y, y\rangle}, \\ \max_{S_{l+1}} \min_{y \in S([N]), y \in S_{l+1}} \frac{\langle \mathcal{T}_N(\mathcal{B}_\mathbb{W}(\mathcal{T}_N(y))), y\rangle}{\langle y, y\rangle}, \end{cases} \\
&= \begin{cases} \min_{S_\ell} \max_{y \in S([N]), y \perp S_\ell} \frac{\int_\mathbb{W} |\widetilde{y}(f)|^2 df}{||y||_2^2}, \\ \max_{S_{l+1}} \min_{y \in S([N]), y \in S_{l+1}} \frac{\int_\mathbb{W} |\widetilde{y}(f)|^2 df}{||y||_2^2}, \end{cases}
\end{aligned} \tag{C.1}$$

where S_ℓ is an ℓ-dimensional subspace of $\ell_2(\mathbb{Z})$, and $\widetilde{y}(f)$ is the DTFT of the sequence y. Noting that all the eigenvectors of $\mathcal{T}_N \mathcal{B}_\mathbb{W} \mathcal{T}_N$ belong to $S([N])$, we restrict to $y \in S([N])$ in the second line.

Lemma C.1. *Consider the bandlimited sequence $g \in \ell_2(\mathbb{Z})$ whose DTFT is given by*

$$\widetilde{g}(f) = \begin{cases} \sqrt{2N} \cos(N\pi f) e^{-j2\pi f \lfloor \frac{N}{2} \rfloor}, & |f| \le \frac{1}{2N}, \\ 0, & \frac{1}{2N} < |f| \le \frac{1}{2}. \end{cases} \tag{C.2}$$

Then $||g||_2^2 = 1$ and $g[n] \ge \frac{1}{\sqrt{2N}}$ for all $n \in [N]$.

Proof of Lemma C.1. . First it is easy to check that $||g||_2^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} |\widetilde{g}(f)|^2 df = 1$. Then computing the inverse DTFT directly yields

$$g[n] = \frac{1}{\sqrt{2N}} \text{sinc}\left(\frac{n - \lfloor \frac{N}{2} \rfloor}{N} - \frac{1}{2}\right) + \frac{1}{\sqrt{2N}} \text{sinc}\left(\frac{n - \lfloor \frac{N}{2} \rfloor}{N} + \frac{1}{2}\right).$$

Let $\xi(t) = \text{sinc}(t - \frac{1}{2}) + \text{sinc}(t + \frac{1}{2})$. Taking the directive of $\xi(t)$, we would find on $[-\frac{1}{2}, \frac{1}{2}]$ that $\xi(t)$ achieves its minimum value of 1 at the points $t = \pm\frac{1}{2}$. Therefore, $g[n] \ge \frac{1}{\sqrt{2N}}$ since $|\frac{n - \lfloor \frac{N}{2} \rfloor}{N}| \le \frac{1}{2}$ for all $n \in [N]$. □

C.3.1 Upper Bound

From equation (C.1), we know that

$$\lambda_{N,\mathbb{W}}^{(\ell)} = \min_{S_\ell} \max_{y \in S([N]), y \perp S_\ell} \frac{\int_{\mathbb{W}} |\widetilde{y}(f)|^2 df}{||y||_2^2}.$$

Therefore, in order to bound the eigenvalues from above, it suffices to pick an appropriate ℓ-dimensional subspace $S_\ell \subset \ell_2(\mathbb{Z})$ and then find a uniform upper bound for the quantity above for all time-limited sequences $y \in S([N])$ orthogonal to S_ℓ.

Consider the bandlimited sequence $g \in \ell_2(\mathbb{Z})$ defined in (C.2). Let $\mathcal{E}_{f_0} : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ denote a modulating operator with $\mathcal{E}_{f_0}(y)[n] := e^{j2\pi f_0 n} y[n]$ for all $n \in \mathbb{Z}$ and $f_0 \in [-\frac{1}{2}, \frac{1}{2}]$. Set

$$L_+ = \{n' \in \mathbb{Z} : -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor,\ (\frac{n'}{N} - \frac{1}{2N}, \frac{n'}{N} + \frac{1}{2N}) \cap \mathbb{W} \neq \emptyset\}$$

and hence $\iota_+ = \#L_+$. Let S_{ι_+} be the ι_+-dimensional subspace of $\ell_2(\mathbb{Z})$ spanned by the functions $\mathcal{E}_{\frac{n'}{N}} g, n' \in L_+$, that is,

$$S_{\iota_+} := \text{span}\left(\{\mathcal{E}_{\frac{n'}{N}} g\}_{n' \in L_+}\right).$$

If the time-limited sequence $y \in S([N])$ is orthogonal to S_{ι_+}, then

$$0 = \langle y, \mathcal{E}_{\frac{n'}{N}} g \rangle = \langle \widetilde{y}, \widetilde{g}(\cdot - \frac{n'}{N}) \rangle = \left(\widetilde{y} \star \widetilde{\overline{g}}\right)\left(\frac{n'}{N}\right) =: g_y[n'],\ n' \in L_+,$$

where $\overline{g} := g^*$ is the complex-conjugate of the sequence g and $\widetilde{\overline{g}}$ is the DTFT of $\overline{g}$.

Now it follows that

$$
\begin{aligned}
\sum_{n'=-\lfloor \frac{N}{2} \rfloor}^{\lfloor \frac{N-1}{2} \rfloor} |g_y[n']|^2 &= \sum_{n' \in L_+^C} |g_y[n']|^2 \\
&= \sum_{n' \in L_+^C} |\int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} \widetilde{y}(f)\widetilde{\overline{g}}(\frac{n'}{N} - f)df|^2 \\
&\leq \sum_{n' \in L_+^C} \left(||g||_2^2 \int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} |\widetilde{y}(f)|^2 df \right) \\
&\leq \int_{f \notin \mathbb{W}} |\widetilde{y}(f)|^2 df \\
&= ||y||_2^2 - \int_{f \in \mathbb{W}} |\widetilde{y}(f)|^2 df,
\end{aligned}
\tag{C.3}
$$

where L_+^C is defined as $L_+^C := \{n' \in \mathbb{Z} : -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor,\ n' \notin L_+\}$, the second line holds because g is bandlimited to $[-\frac{1}{2N}, \frac{1}{2N}]$, the third line follows from the Cauchy-Schwarz inequality, and the fourth line holds because $||g||_2 = 1$ and by construction, the set $\cup_{n' \in L_+}[\frac{n'}{N} - \frac{1}{2N}, \frac{n'}{N} + \frac{1}{2N}]$ covers the intervals $\mathbb{W}$ completely. On the other hand, let $y \odot \overline{g}$ denote the pointwise product between y and $\overline{g}$, that is $(y \odot \overline{g})[n] = y[n]\overline{g}[n]$. Note that $y \odot \overline{g}$ has the same support in time as y, namely $[N]$, and $\{\frac{1}{\sqrt{N}} \boldsymbol{e}_{\frac{n'}{N}}, -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor\}$ forms an orthobasis (normalized DFT basis) for $\mathbb{C}^N$. We can rewrite $g_y[n'] = \boldsymbol{e}_{\frac{n'}{N}}^H (y \odot \overline{g})$, which can be viewed as the DFT of $y \odot \overline{g}$. Therefore, using Parseval's theorem, we acquire

$$
\sum_{n'=-\lfloor \frac{N}{2} \rfloor}^{\lfloor \frac{N-1}{2} \rfloor} |g_y[n']|^2 = N||y \odot \overline{g}||_2^2 \geq \frac{1}{2}||y||_2^2
$$

since by hypothesis, $g[n] \geq \frac{1}{\sqrt{2N}}$ for all $n \in [N]$. Now, combining the above lower bound on the energy of the sequence g_y and the upper bound in (C.3), we observe that

$$
\frac{1}{2}||y||_2^2 \leq ||y||_2^2 - \int_{f \in \mathbb{W}} |\widetilde{y}(f)|^2 df,
$$

and therefore,

$$
\lambda_{N,\mathbb{W}}^{(\iota_+)} \leq \frac{\int_{\mathbb{W}} |\widetilde{y}(f)|^2 df}{||y||_2^2} \leq \frac{1}{2}.
$$

C.3.2 Lower Bound

In the other direction, consider the minimax representation

$$\lambda_{N,\mathbb{W}}^{(\ell)} = \max_{S_{l+1}} \min_{y \in S([N]), y \in S_{l+1}} \frac{\int_{\mathbb{W}} |\widetilde{y}(f)|^2 df}{||y||_2^2}.$$

In order to find a lower bound for the eigenvalues, it suffices to pick an appropriate $(l+1)$-dimensional subspace $S_{l+1} \subset \ell_2(\mathbb{Z})$ and then find a uniform lower bound for the quantity above for all time-limited sequences $y \in S([N])$ inside S_{l+1}. With g as defined in (C.2), let the time-limited sequence $h \in \ell_2([N])$ be such that $h[n] = 1/\overline{g}[n]$ for all $n \in [N]$. We set

$$L_- := \{n' \in \mathbb{Z} : -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor,\ (\frac{n'}{N} - \frac{1}{2N}, \frac{n'}{N} + \frac{1}{2N}) \subset \mathbb{W}\},$$

and hence $\iota_- = \#L_-$. Let S_{ι_-} be the ι_--dimensional subspace of $\ell_2(\mathbb{Z})$ spanned by the functions $\mathcal{E}_{\frac{n'}{N}} h, n' \in L_-$, that is,

$$S_{\iota_-} := \operatorname{span}\left(\{\mathcal{E}_{\frac{n'}{N}} h\}_{n' \in L_-} \right).$$

Suppose $y \in S_{\iota_-}$ (and hence $y \in \ell_2([N])$). Then we may write

$$y = \sum_{n' \in L_-} b_{n'} \mathcal{E}_{\frac{n'}{N}} h$$

for some coefficients $b_{n'}$. Moreover,

$$y \odot \overline{g} = \sum_{n' \in L_-} b_{n'} \boldsymbol{e}_{\frac{n'}{N}}.$$

Noting that $\{\frac{1}{\sqrt{N}} \boldsymbol{e}_{\frac{n'}{N}}, -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor\}$ forms an orthobasis for $\mathbb{C}^N$, we obtain

$$\sum_{n' \in L_-} |b_{n'}|^2 = N||y \odot \overline{g}||_2^2 = N \sum_{n=0}^{N-1} |y[n] \odot \overline{g}[n]|^2 \geq \frac{1}{2} \sum_{n=0}^{N-1} |y[n]|^2 = \frac{1}{2}||y||_2^2$$

since by definition, $g[n] \geq \frac{1}{\sqrt{2N}}$ for all $n \in [N]$. On the other hand,

$$b_{n'} = \sum_{n=0}^{N-1} \overline{g}[n]y[n]e^{-j2\pi\frac{n'}{N}n} = \langle y, \mathcal{E}_{\frac{n'}{N}} g\rangle.$$

Now using the same procedure as in (C.3), one has

$$\begin{aligned}
\sum_{n'\in L_-} |b_{n'}|^2 &= \sum_{n'\in L_-} |\langle y, \mathcal{E}_{\frac{n'}{N}} g\rangle|^2 \\
&= \sum_{n'\in L_-} |\int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} \widetilde{y}(f)\widetilde{\overline{g}}(\frac{n'}{N} - f)df|^2 \\
&\leq \sum_{n'\in L_-} \left(||g||_2^2 \int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} |\widetilde{y}(f)|^2 df \right) \\
&\leq \int_{f\in\mathbb{W}} |\widetilde{y}(f)|^2 df,
\end{aligned}$$

where the last line holds since by construction, the set $\cup_{n'\in L_i}[\frac{n'}{N} - \frac{1}{2N}, \frac{n'}{N} + \frac{1}{2N}]$ is a subset of the intervals $\mathbb{W}$. Altogether, we then conclude that for any $y \in S_{\iota_-}$ (and hence $y \in S([N])$),

$$\frac{1}{2}||y||_2^2 \leq \int_{f\in\mathbb{W}} |\widetilde{y}(f)|^2 df.$$

And hence

$$\lambda_{N,\mathbb{W}}^{(\iota_- - 1)} \geq \frac{\int_{f\in\mathbb{W}} |\widetilde{y}(f)|^2 df}{||y||_2^2} \geq \frac{1}{2}.$$

C.4 Proof of Theorem 5.3

Proof of eigenvalues that cluster near zero

Since $\boldsymbol{B}_{N,\mathbb{W}} = \sum_{i=0}^{J-1} \boldsymbol{E}_{f_i} \boldsymbol{B}_{N,W_i} \boldsymbol{E}_{f_i}^H$, according to [67] (see pp. 181), the following holds

$$\lambda_{N,\mathbb{W}}^{(\ell)} \leq \sum_{i=0}^{J-1} \lambda_{N,W_i}^{(\ell_i)}$$

for all $\ell_i \in [N], i \in [J]$ and $\ell = \sum_{i=0}^{J-1} l_i \in [N]$.

Fix $\epsilon \in (0, \frac{1}{|\mathbb{W}|} - 1)$. For each $i \in [J]$, let $N_1(W_i, \epsilon)$, $C_3(W_i, \epsilon)$ and $C_4(W_i, \epsilon)$ be the constants specified in Lemma 2.1 with respect to W_i and ϵ. If we let

$$\overline{N}_1(\mathbb{W}, \epsilon) = \max\{N_1(W_i, \epsilon),\ \forall\ i \in [J]\},$$

then we have

$$\lambda_{N,W_i}^{(\ell_i)} \leq C_3(W_i, \epsilon) e^{-C_4(W_i,\epsilon)N},\ \forall\ l_i \geq \lceil 2NW_i(1+\epsilon) \rceil, i \in [J]$$

for all $N \geq \overline{N}_1(\mathbb{W}, \epsilon)$. Hence, by choosing $\ell_i \geq \lceil 2NW_i(1+\epsilon) \rceil,\ \forall\ i \in [J]$, we have

$$\lambda_{N,\mathbb{W}}^{(\ell)} \leq \sum_{i=0}^{J-1} C_3(W_i, \epsilon) e^{-C_4(W_i,\epsilon)N} \leq \overline{C}_3(\mathbb{W}, \epsilon) e^{-\overline{C}_4(\mathbb{W},\epsilon)N},$$

for all $N \geq \overline{N}_1(\mathbb{W}, \epsilon)$ and $\ell \geq \sum_i \lceil 2NW_i(1+\epsilon) \rceil$,where $\overline{C}_3(\mathbb{W}, \epsilon) = J \max\{C_3(W_i, \epsilon),\ \forall\ i \in [J]\}$ and $\overline{C}_4(\mathbb{W}, \epsilon) = \min\{C_4(W_i, \epsilon),\ \forall\ i \in [J]\}$.

ε-pseudo eigenvalue and eigenvectors

Definition C.1. *(ε-pseudo eigenvalue and eigenvector [105]) Let $\boldsymbol{X} \in \mathbb{C}^{N \times N}$ be any matrix and $\boldsymbol{u} \in \mathbb{C}^N$ be any vector with unit ℓ_2-norm. Given $\varepsilon > 0$, the number $\lambda \in \mathbb{C}$ and vector $\boldsymbol{u} \in \mathbb{C}^N$ are an ε-pseudo eigenpair of $\boldsymbol{X}$ if the following condition is satisfied:*

$$||(\boldsymbol{X} - \lambda \boldsymbol{I})\boldsymbol{u}||_2^2 \leq \varepsilon.$$

Lemma C.2. *Suppose $\mathbb{W}$ is a fixed finite union of J pairwise disjoint intervals as defined in (2.6). Fix $\epsilon \in (0, 1)$. For each $i \in [J]$, let $N_0(W_i, \epsilon)$ be the constant specified in Lemma 2.1 with respect to W_i and ϵ and let $\widetilde{N}_0(\mathbb{W}, \epsilon) = \max\{N_0(W_i, \epsilon),\ \forall\ i \in [J]\}$. Then for all $\ell_i \leq 2NW_i(1-\epsilon), i \in [J]$ and $N > \widetilde{N}_0(\mathbb{W}, \epsilon)$, $(\lambda_{N,W_i}^{(\ell_i)}, \boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)})$ is an ε-pseudo eigenpair of $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ with $\varepsilon \leq 2C_1(W_i, \epsilon) e^{-C_2(W_i,\epsilon)N}$, or in detail*

$$\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)}))) = \lambda_{N,W_i}^{(\ell_i)} \boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)} + \boldsymbol{o}_i^{(\ell_i)},$$

where $\boldsymbol{o}_i^{(\ell_i)} = \mathcal{I}_N(\mathcal{B}_{\mathbb{W}\setminus[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)})))$ *and* $||\boldsymbol{o}_i^{(\ell_i)}||_2^2 \leq 2C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N}$. *Here* $\mathbb{W}\setminus[f_i-W_i,f_i+W_i] = \bigcup_{i'\neq i}[f_{i'}-W_{i'},f_{i'}+W_{i'}]$ *means the set difference between* $\mathbb{W}$ *and* $[f_i-W_i,f_i+W_i]$, *and* $C_1(W_i,\epsilon)$ *and* $C_2(W_i,\epsilon)$ *are the constants specified in Lemma 2.1 corresponding to* W_i *and* ϵ *for all* $i\in[J]$.

Proof of Lemma C.2. According to the definition of the operator $\mathcal{I}_N\mathcal{B}_{\mathbb{W}}\mathcal{I}_N^*$,

$$
\begin{aligned}
&\left(\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)})))\right)[m]\\
=&\sum_{n=0}^{N-1}\sum_{i'=0}^{J-1} e^{j2\pi f_{i'}(m-n)}\frac{\sin(2\pi W_{i'}(m-n))}{\pi(m-n)}e^{j2\pi f_i n}\boldsymbol{s}_{N,W_i}^{(\ell_i)}[n]\\
=&e^{j2\pi f_i m}\lambda_{N,W_i}^{(\ell_i)}\boldsymbol{s}_{N,W_i}^{(\ell_i)}[m] + \sum_{n=0}^{N-1}\sum_{i'=0,i'\neq i}^{J-1} e^{j2\pi f_{i'}(m-n)}\frac{\sin(2\pi W_{i'}(m-n))}{\pi(m-n)}e^{j2\pi f_i n}\boldsymbol{s}_{N,W_i}^{(\ell_i)}[n]\\
=&e^{j2\pi f_i m}\lambda_{N,W_i}^{(\ell_i)}\boldsymbol{s}_{N,W_i}^{(\ell_i)}[m] + \mathcal{I}_N(\mathcal{B}_{\mathbb{W}\setminus[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)})))[m].
\end{aligned}
$$

In what follows, we will bound the energy of $\boldsymbol{o}_i^{(\ell_i)} = \mathcal{I}_N(\mathcal{B}_{\mathbb{W}\setminus[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)})))$ as

$$
\begin{aligned}
||\boldsymbol{o}_i^{(\ell_i)}||_2^2 &= ||\mathcal{I}_N(\mathcal{B}_{\mathbb{W}\setminus[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)})))||_2^2\\
&\leq ||\mathcal{B}_{\mathbb{W}\setminus[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)}))||_2^2\\
&\leq ||\mathcal{B}_{[-\frac{1}{2},\frac{1}{2}]\setminus[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)}))||_2^2\\
&= ||\boldsymbol{s}_{N,W_i}^{(\ell_i)}||_2^2 - ||\mathcal{B}_{[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)}))||_2^2\\
&\leq ||\boldsymbol{s}_{N,W_i}^{(\ell_i)}||_2^2 - ||\mathcal{I}_N(\mathcal{B}_{[f_i-W_i,f_i+W_i]}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)})))||_2^2\\
&\leq 1-(\lambda_{N,W_i}^{(\ell_i)})^2 \leq 1-(1-C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N})^2\\
&= 2C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N} - (C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N})^2 \leq 2C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N}
\end{aligned}
$$

for all $\ell_i \leq \lfloor 2NW_i(1-\epsilon)\rfloor, i\in[J]$ and $N\geq \widetilde{N}_0(\mathbb{W},\epsilon)$. Here the second inequality in the sixth line follows simply from Lemma 2.1 since $\widetilde{N}_0(\mathbb{W},\epsilon)\geq N_0(W_i,\epsilon)$. □

Using this result, we now show the first $\approx N|\mathbb{W}|$ eigenvalues of $\mathcal{I}_N\mathcal{B}_{\mathbb{W}}\mathcal{I}_N^*$ are close to 1.

Proof of eigenvalues that cluster near one

The main idea is to guarantee that the sum of the first $\approx N|\mathbb{W}|$ eigenvalues is sufficiently close $N|\mathbb{W}|$. Then we conclude that the first $\approx N|\mathbb{W}|$ eigenvalues cluster near one by applying the fact that the eigenvalues are upper bounded by 1. First we state the following useful results.

Lemma C.3. *([32] Lemma 5.1) Fix $\epsilon \in (0,1)$. Let $k_i = \lfloor 2NW_i(1-\epsilon) \rfloor,\ \forall\ i \in [J]$, and let $\boldsymbol{\Psi}$ be the dictionary as defined in (5.1). Then for any pair of distinct columns $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ in $\boldsymbol{\Psi}$, we have*

$$|\langle \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \rangle| \leq 3\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)} e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2} N} \tag{C.4}$$

and

$$\left\| \boldsymbol{\Psi}^H \boldsymbol{\Psi} \right\|_2 \leq 1 + 3N\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)} e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}$$

if $N \geq \widetilde{N}_0(\mathbb{W},\epsilon)$, where

$$\widetilde{C}_1(\mathbb{W},\epsilon) = \max\{C_1(W_i,\epsilon),\ \forall\ i \in [J]\}, \widetilde{C}_2(\mathbb{W},\epsilon) = \min\{C_2(W_i,\epsilon),\ \forall\ i \in [J]\}.$$

Here $||\boldsymbol{\Psi}^H\boldsymbol{\Psi}||_2$ is the spectral norm (or largest singular value) of $\boldsymbol{\Psi}^H\boldsymbol{\Psi}$.

Lemma C.4. *([67]) Let $\boldsymbol{X} \in \mathbb{C}^{N\times N}$ be a Hermitian matrix, and let $\lambda_0(\boldsymbol{X}), \lambda_1(\boldsymbol{X}), \ldots, \lambda_{N-1}(\boldsymbol{X})$ be its eigenvalues arranged in decreasing order. Then,*

$$\lambda_0(\boldsymbol{X}) + \lambda_1(\boldsymbol{X}) + \ldots + \lambda_{r-1}(\boldsymbol{X}) = \max_{\boldsymbol{U}\in\mathbb{C}^{N\times r}, \boldsymbol{U}^H\boldsymbol{U}=\boldsymbol{I}_r} \text{trace}(\boldsymbol{U}^H\boldsymbol{X}\boldsymbol{U}),$$

where $\boldsymbol{I}_r$ is the $r \times r$ identity matrix with $1 \leq r \leq N$.

Based on this result, we propose the following generalized result concerning the sum of the first r eigenvalues.

Lemma C.5. *Let $\boldsymbol{X} \in \mathbb{C}^{N\times N}$ be a positive-semidefinite (PSD) matrix, and let $\lambda_i(\boldsymbol{X}), 0 \leq i \leq N-1$ be its eigenvalues arranged in decreasing order. Then, for any matrix $\boldsymbol{M} \in \boldsymbol{C}^{N\times r}, 1 \leq r \leq N$, the following inequality holds*

$$\lambda_0(\boldsymbol{X}) + \lambda_1(\boldsymbol{X}) + \ldots + \lambda_{r-1}(\boldsymbol{X}) \geq \text{trace}(\boldsymbol{M}^H \boldsymbol{X} \boldsymbol{M})/\|\boldsymbol{M}^H \boldsymbol{M}\|_2.$$

Proof of Lemma C.5. Let $\sigma_0(\boldsymbol{M}), \ldots, \sigma_{r-1}(\boldsymbol{M})$ denote the decreasing singular values of the matrix $\boldsymbol{M}$. Denote $\boldsymbol{M} = \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^H$ as the truncated SVD of $\boldsymbol{M}$, where $\boldsymbol{\Sigma}_r$ is an $r \times r$ diagonal matrix with $\sigma_0(\boldsymbol{M}), \ldots, \sigma_{r-1}(\boldsymbol{M})$ along its diagonal.

Now applying Lemma C.4, we obtain

$$\begin{aligned}
\sum_{l=0}^{r-1} \lambda_\ell(\boldsymbol{X}) &\geq \text{trace}(\boldsymbol{U}_r^H \boldsymbol{X} \boldsymbol{U}_r) \\
&\geq \text{trace}(\boldsymbol{\Sigma}_r \boldsymbol{U}_r^H \boldsymbol{X} \boldsymbol{U}_r \boldsymbol{\Sigma}_r)/(\sigma_0(\boldsymbol{M}))^2 \\
&= \text{trace}(\boldsymbol{V_r} \boldsymbol{\Sigma_r} \boldsymbol{U}_r^H \boldsymbol{X} \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^H)/\|\boldsymbol{M}^H \boldsymbol{M}\|_2 \\
&= \text{trace}(\boldsymbol{M}^H \boldsymbol{X} \boldsymbol{M})/\|\boldsymbol{M}^H \boldsymbol{M}\|_2,
\end{aligned}$$

where the first line follows directly from Lemma C.4, the second line is obtained because $\boldsymbol{U}_r^H \boldsymbol{X} \boldsymbol{U}_r$ is PSD and hence its main diagonal elements are non-negative, and the third line follows because $\boldsymbol{V}_r$ is an orthobasis and $(\sigma_0(\boldsymbol{M}))^2 = \|\boldsymbol{M}^H \boldsymbol{M}\|_2$. □

We are now ready to prove the main part. Fix $\epsilon \in (0, 1)$. Let $k_i = \lfloor 2NW_i(1-\epsilon) \rfloor, \forall i \in [J]$, and let $\boldsymbol{\Psi}$ be the dictionary as defined in (5.1). We have

$$
\begin{aligned}
&\sum_{\ell=0}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon)\rfloor} \lambda_{N,\mathbb{W}}^{(\ell)} \\
&\geq \operatorname{trace}\left(\boldsymbol{\Psi}^H \boldsymbol{B}_{N,\mathbb{W}} \boldsymbol{\Psi}\right) / \left\|\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right\|_2 \\
&= \left(\sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon)\rfloor} \left((\boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)})))\right)\right) / \left\|\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right\|_2 \\
&= \left(\sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon)\rfloor} \left(\left(\boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)}\right)^H \left(\lambda_{N,W_i}^{(\ell_i)} \boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)} + \boldsymbol{o}_i^{(\ell_i)}\right)\right)\right) / \left\|\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right\|_2 \\
&\geq \left(\sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon)\rfloor} \left(\lambda_{N,W_i}^{(\ell_i)} - \|\boldsymbol{o}_i^{(\ell_i)}\|_2\right)\right) / \left\|\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right\|_2 \\
&\geq \frac{\left(\sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon)\rfloor} \left(1 - C_1(W_i,\epsilon) e^{-C_2(W_i,\epsilon)N} - \sqrt{2}\sqrt{C_1(W_i,\epsilon)} e^{-\frac{C_2(W_i,\epsilon)}{2}N}\right)\right)}{\left(1 + 3N\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)} e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}\right)} \\
&\geq \frac{\left(\sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon)\rfloor} \left(1 - \widetilde{C}_1(\mathbb{W},\epsilon) e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N} - \sqrt{2}\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)} e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}\right)\right)}{\left(1 + 3N\sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)} e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}\right)} \\
&\geq \frac{J + \sum_i \lfloor 2NW_i(1-\epsilon)\rfloor - 3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}}{1 + 3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}} \\
&= \frac{\left(J + \sum_i \lfloor 2NW_i(1-\epsilon)\rfloor - 3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}\right)\left(1 - 3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}\right)}{\left(1 + 3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}\right)\left(1 - 3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}\right)} \\
&\geq \frac{J + \sum_i \lfloor 2NW_i(1-\epsilon)\rfloor - 6N^2 C_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N} + \left(3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}\right)^2}{1 - \left(3NC_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}}\right)^2} \\
&\geq J + \sum_i \lfloor 2NW_i(1-\epsilon)\rfloor - 6N^2 C_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}N}
\end{aligned}
$$

for all $N \geq \max\{\widetilde{N}_0(\mathbb{W},\epsilon), N'(\mathbb{W},\epsilon)\}$, where

$$
N'(\mathbb{W},\epsilon) = \max\{(\frac{4}{C_2(\mathbb{W},\epsilon)})^2,\ \frac{4}{C_2(\mathbb{W},\epsilon)} \log(3C_5(\mathbb{W},\epsilon))\}
$$

is the constant such that $3NC_5(\mathbb{W},\epsilon)e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}} < 1$ for all $N \geq N'(\mathbb{W},\epsilon)$.[21] Here the first line follows directly from Lemma C.5, the second line follows because trace $\left(\boldsymbol{\Psi}^H \boldsymbol{B}_{N,\mathbb{W}} \boldsymbol{\Psi}\right) =$ trace $\left(\sum_{i=0}^{J-1} \boldsymbol{\Psi}_i^H \boldsymbol{B}_{N,\mathbb{W}} \boldsymbol{\Psi}_i\right)$ and $\boldsymbol{B}_{N,\mathbb{W}}$ is equivalent to $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$, the third line follows from Lemma C.2, the fourth line follows from the Cauchy-Schwarz inequality which indicates that $|(\boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \boldsymbol{o}_i^{(\ell_i)}| \leq ||\boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)}||_2 ||\boldsymbol{o}_i^{(\ell_i)}||_2 = ||\boldsymbol{o}_i^{(\ell_i)}||_2$, the fifth line follows from Lemmas 2.1, C.2 and C.3, the seventh line follows by setting $C_5(\mathbb{W},\epsilon) = \max\{\widetilde{C}_1(\mathbb{W},\epsilon), \sqrt{\widetilde{C}_1(\mathbb{W},\epsilon)}\}$, the ninth line follows because $J + \sum_i \lfloor 2NW(1-\epsilon) \rfloor \leq N$, and the last line follows because by assumption $3NC_5(\mathbb{W},\epsilon)e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}} < 1$.

By noting that $0 < \lambda_{N,\mathbb{W}}^{(N-1)} \leq \lambda_{N,\mathbb{W}}^{(0)} < 1$ from Lemma 5.1, we acquire

$$\begin{aligned}
\lambda_{N,\mathbb{W}}^{(\ell)} &= \left(\sum_{\ell'=0}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(\ell')} \right) - \left(\sum_{\ell'=0,\ell' \neq l}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(\ell')} \right) \\
&\geq \left(\sum_{\ell'=0}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(\ell')} \right) - \left(J - 1 + \sum_i \lfloor 2NW_i(1-\epsilon) \rfloor \right) \\
&\geq 1 - 6N^2 C_5(\mathbb{W},\epsilon) e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2} N}
\end{aligned}$$

for all $\ell \leq J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor$, where the second line follows by setting $\lambda_{N,\mathbb{W}}^{(\ell')},\ \ell' \neq l$ to 1. Fix $\mathbb{W}$ and ϵ. It is always possible to find a constant N' such that $3NC_5(\mathbb{W},\epsilon)e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}} < 1$ for all $N \geq N'$. Now, for convenience, we set $\overline{C}_1(\mathbb{W},\epsilon) = 6C_5(\mathbb{W},\epsilon)$, $\overline{C}_2(\mathbb{W},\epsilon) = \frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2}$, and $\overline{N}_0(\mathbb{W},\epsilon) = \max\{\widetilde{N}_0(\mathbb{W},\epsilon), N'\}$. This completes the proof of Theorem 5.3.

C.5 Proof of Theorem 5.4

First denote the eigen-decomposition of $\boldsymbol{B}_{N,\mathbb{W}}$ as

$$\boldsymbol{B}_{N,\mathbb{W}} = \boldsymbol{U}_{N,\mathbb{W}} \boldsymbol{\Lambda}_{N,\mathbb{W}} \boldsymbol{U}_{N,\mathbb{W}}^H,$$

[21] This can be verified as $3NC_5(\mathbb{W},\epsilon)e^{-\frac{\widetilde{C}_2(\mathbb{W},\epsilon)N}{2}} = 3C_5(\mathbb{W},\epsilon)e^{-N(\frac{\widetilde{C}_2(\mathbb{W},\epsilon)}{2} - \frac{\log N}{N})} \leq 3C_5(\mathbb{W},\epsilon)e^{-N\frac{\widetilde{C}_2\mathbb{W},\epsilon)}{4}} \leq 1$ for all $N \geq \max\{(\frac{4}{C_2(\mathbb{W},\epsilon)})^2,\ \frac{4}{C_2(\mathbb{W},\epsilon)} \log(3C_5(\mathbb{W},\epsilon))\}$. Here the first inequality follows because $\frac{\log N}{N} \leq \frac{1}{N^{1/2}} \leq \frac{C_2(\mathbb{W},\epsilon)}{4}$ for all $N \geq (\frac{4}{C_2(\mathbb{W},\epsilon)})^2$.

where $\boldsymbol{\Lambda}_{N,\mathbb{W}}$ is an $N \times N$ diagonal matrix whose diagonal elements are the eigenvalues $\lambda^{(0)}_{N,\mathbb{W}}, \lambda^{(1)}_{N,\mathbb{W}}, \ldots, \lambda^{(N-1)}_{N,\mathbb{W}}$ and $\boldsymbol{U}_{N,\mathbb{W}}$ is a square $(N \times N)$ matrix defined by

$$\boldsymbol{U}_{N,\mathbb{W}} := [\boldsymbol{u}^{(0)}_{N,\mathbb{W}}\ \boldsymbol{u}^{(1)}_{N,\mathbb{W}}\ \ldots\ \boldsymbol{u}^{(N-1)}_{N,\mathbb{W}}].$$

Also let $\boldsymbol{a} = \boldsymbol{U}^H_{N,\mathbb{W}}\boldsymbol{\psi}$ be the coefficients of $\boldsymbol{\psi}$ represented by $\boldsymbol{U}_{N,\mathbb{W}}$.

Fix $\epsilon \in (0, \min\{1, \frac{1}{|\mathbb{W}|} - 1\})$. Suppose $\boldsymbol{\psi}$ is a column of $\boldsymbol{\Psi}_i$ for some particular $i \in [J]$. Now from Lemma C.2, we have

$$\boldsymbol{B}_{N,\mathbb{W}}\boldsymbol{\psi} = \lambda^{(\ell_i)}_{N,W_i}\boldsymbol{\psi} + \boldsymbol{o}^{(\ell_i)}_i$$

for some $\ell_i \leq \lfloor 2NW_i(1-\epsilon) \rfloor$.

Plugging the eigen-decomposition of the matrix $\boldsymbol{U}_{N,\mathbb{W}}$ into the above equation, we require

$$\boldsymbol{\Lambda}_{N,\mathbb{W}}\boldsymbol{a} = \lambda^{(\ell_i)}_{N,W_i}\boldsymbol{a} + \widehat{\boldsymbol{o}}^{(\ell_i)}_i,$$

where $\widehat{\boldsymbol{o}}^{(\ell_i)}_i = \boldsymbol{U}^H_{N,\mathbb{W}}\boldsymbol{o}^{(\ell_i)}_i$. The elementary form of the above equation is

$$\lambda^{(m)}_{N,\mathbb{W}}\boldsymbol{a}[m] = \lambda^{(\ell_i)}_{N,W_i}\boldsymbol{a}[m] + \widehat{\boldsymbol{o}}^{(\ell_i)}_i[m]$$

for all $m \in [N]$.

Now we have

$$\begin{aligned}
||\boldsymbol{\psi} - \boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{\psi}||^2_2 = \sum_{m=\sum_i \lceil 2NW_i(1+\epsilon)\rceil}^{N-1} |\boldsymbol{a}[m]|^2 &= \sum_{m=\sum_i \lceil 2NW_i(1+\epsilon)\rceil}^{N-1} \frac{\left|\widehat{\boldsymbol{o}}^{(\ell_i)}_i[m]\right|^2}{\left|\lambda^{(\ell_i)}_{N,W_i} - \lambda^{(m)}_{N,\mathbb{W}}\right|^2} \\
&\leq \frac{\sum_{m=\sum_i \lceil 2NW_i(1+\epsilon)\rceil}^{N-1} \left|\widehat{\boldsymbol{o}}^{(\ell_i)}_i[m]\right|^2}{\left(1 - \widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N} - \overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}\right)^2} \\
&\leq \frac{||\boldsymbol{o}^{(\ell_i)}_i||^2}{\left(1 - \widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N} - \overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}\right)^2} \\
&\leq \frac{2\widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N}}{\left(1 - \widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N} - \overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}\right)^2}
\end{aligned} \tag{C.5}$$

for all $N \geq \max\{\overline{N}_0(\mathbb{W},\epsilon), \overline{N}_1(\mathbb{W},\epsilon)\}$, where the second line follows by bounding the $\lambda_{N,W_i}^{(\ell_i)}$ term using $1 - C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N}$ (which is not less than $1 - \widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N}$) from Lemma 2.1 and bounding the $\lambda_{N,\mathbb{W}}^{(m)}$ terms using Theorem 5.3, and the fourth line follows because $||\boldsymbol{o}_i^{(\ell_i)}||^2 \leq 2C_1(W_i,\epsilon)e^{-C_2(W_i,\epsilon)N} \leq 2\widetilde{C}_1(\mathbb{W},\epsilon)e^{-\widetilde{C}_2(\mathbb{W},\epsilon)N}$.

The following general result will help in extending (C.5) to an angle between the subspaces.

Lemma C.6. *Let $\mathcal{S}_{\boldsymbol{U}}$ and $\mathcal{S}_{\boldsymbol{V}}$ be the subspaces spanned by the columns of the matrices $\boldsymbol{U} \in \mathbb{C}^{N\times q}$ and $\boldsymbol{V} \in \mathbb{C}^{N\times r}$, respectively. Here $r \leq q \leq N$. Suppose each column of $\boldsymbol{V}$ is normalized so that $\|\boldsymbol{v}_\ell\|_2 = 1$ and is close to $\mathcal{S}_{\boldsymbol{U}}$ such that for some δ_1, $\|\boldsymbol{v}_\ell - \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_\ell\|_2^2 \leq \delta_1$ for all $\ell \in [r]$. Furthermore, suppose the columns of $\boldsymbol{V}$ are approximately orthogonal to each other such that for some δ_2, $|\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell\rangle| \leq \delta_2$ for all $k \neq l$. Then we have*

$$\cos(\Theta_{\mathcal{S}_{\boldsymbol{U}}\mathcal{S}_{\boldsymbol{V}}}) \geq \sqrt{\frac{1 - \delta_1 - N\left(\delta_2 + \sqrt{\delta_1}\right)}{1 + N\delta_2}}.$$

Proof of Lemma C.6. Any $\boldsymbol{v} \in \mathcal{S}_{\boldsymbol{V}}$ can be written as a linear combination of $\boldsymbol{v}_\ell$ in the form $\boldsymbol{v} = \sum_\ell \alpha_\ell \boldsymbol{v}_\ell$. We first bound the ℓ_2 norm of $\boldsymbol{v}$ by

$$\begin{aligned}
\|\boldsymbol{v}\|_2^2 &= \|\sum_{\ell=0}^{r-1} \alpha_\ell \boldsymbol{v}_\ell\|_2^2 \\
&= \sum_{\ell=0}^{r-1} |\alpha_\ell|^2 \|\boldsymbol{v}_\ell\|_2^2 + \sum_{\ell=0}^{r-1}\sum_{k=0,k\neq l}^{r-1} \langle \alpha_\ell \boldsymbol{v}_\ell, \alpha_k \boldsymbol{v}_k\rangle \\
&\leq \sum_{\ell=0}^{r-1} |\alpha_\ell|^2 + \sum_{\ell=0}^{r-1}\sum_{k=0,k\neq l}^{r-1} |\alpha_\ell||\alpha_k|\delta_2 \\
&\leq \sum_{\ell=0}^{r-1} |\alpha_\ell|^2 + \sum_{\ell=0}^{r-1}\sum_{k=0,k\neq l}^{r-1} \frac{|\alpha_\ell|^2 + |\alpha_k|^2}{2}\delta_2 \\
&= \left(\sum_{\ell=0}^{r-1} |\alpha_\ell|^2\right)(1 + (r-1)\delta_2) \leq \left(\sum_{\ell=0}^{r-1} |\alpha_\ell|^2\right)(1 + N\delta_2),
\end{aligned}$$

where the third line follows from the hypothesis that $|\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell \rangle| \leq \delta_2$ for all $k \neq l$. Similarly,

$$\begin{aligned}
\|\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}\|_2^2 &= \|\sum_{\ell=0}^{r-1} \boldsymbol{P}_{\boldsymbol{U}}\left(\alpha_\ell \boldsymbol{v}_\ell\right)\|_2^2 = \sum_{\ell=0}^{r-1} |\alpha_\ell|^2 \|\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_\ell\|_2^2 + \sum_{\ell=0}^{r-1}\sum_{k=0,k\neq l}^{r-1} \langle \alpha_\ell \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_\ell, \alpha_k \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_k\rangle \\
&= \sum_{\ell=0}^{r-1} |\alpha_\ell|^2 \|\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_\ell\|_2^2 + \sum_{\ell=0}^{r-1}\sum_{k=0,k\neq l}^{r-1} \langle \alpha_\ell \boldsymbol{v}_\ell, \alpha_k \left(\boldsymbol{v}_k - (\boldsymbol{v}_k - \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_k)\right)\rangle \\
&\geq \sum_{\ell=0}^{r-1} |\alpha_\ell|^2 (1-\delta_1) - \sum_{\ell=0}^{r-1}\sum_{k=0,k\neq l}^{r-1} |\alpha_\ell||\alpha_k| \left(\delta_2 + \sqrt{\delta_1}\right) \\
&= \left(\sum_{\ell=0}^{r-1} |\alpha_\ell|^2\right)\left(1 - \delta_1 - (r-1)\left(\delta_2 + \sqrt{\delta_1}\right)\right) \geq \left(\sum_{\ell=0}^{r-1} |\alpha_\ell|^2\right)\left(1 - \delta_1 - N\left(\delta_2 + \sqrt{\delta_1}\right)\right),
\end{aligned}$$

where the fourth line follows because $\langle \boldsymbol{v}_\ell, \boldsymbol{v}_k - \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_k\rangle \leq \|\boldsymbol{v}_\ell\|_2 \|\boldsymbol{v}_k - \boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}_k\|_2 \leq \sqrt{\delta_1}$ and $|\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell\rangle| \leq \delta_2$ for all $k \neq l$.

Therefore, for any non-zero vector $\boldsymbol{v} \in \mathcal{S}_{\boldsymbol{V}}$ we have

$$\frac{\|\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{v}\|_2^2}{\|\boldsymbol{v}\|_2^2} \geq \frac{1 - \delta_1 - N\left(\delta_2 + \sqrt{\delta_1}\right)}{1 + N\delta_2}.$$

□

Finally, (5.4) follows from Lemma C.6 by replacing $\boldsymbol{U}$ with $\boldsymbol{\Phi}$ and $\boldsymbol{V}$ with $\boldsymbol{\Psi}$, and assigning δ_1 with the upper bound in (C.5) and δ_2 with the upper bound in (C.4).

C.6 Proof of Theorem 5.5

For each $i \in [J]$, define $\overline{\boldsymbol{\Psi}}_i = [\boldsymbol{E}_{f_i} \boldsymbol{S}_{N,W_i} \sqrt{\boldsymbol{\Lambda}_{N,W_i}}]_{k_i}$ for some given $k_i \in \{1, 2, \ldots, N\}$. We construct the scaled multiband modulated DPSS matrix $\overline{\boldsymbol{\Psi}}$ by[22]

$$\overline{\boldsymbol{\Psi}} := [\overline{\boldsymbol{\Psi}}_0 \ \overline{\boldsymbol{\Psi}}_1 \ \cdots \ \overline{\boldsymbol{\Psi}}_{J-1}]. \tag{C.6}$$

The main idea is to bound $\left\|\boldsymbol{P}_{\boldsymbol{\Psi}} \boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2$ using $\left\|\overline{\boldsymbol{\Psi}}\ \overline{\boldsymbol{\Psi}}^H \boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2$. In order to use this argument, we first give out some useful results.

[22] Hogan and Lakey [64] considered the scaled and shifted Prolate Spheroidal Wave Fuctions (PSWF's) and provided conditions on a shift parameter such that the scaled and shifted PSWF's form a frame or a Riesz basis for the Paley-Wiener space.

Lemma C.7. *Suppose $\overline{\boldsymbol{\Psi}}$ is the matrix defined in (C.6) with some given $k_i \in \{1, 2, \ldots, N\}, \forall i \in [J]$. Then*

$$\left\|\overline{\boldsymbol{\Psi}}\right\|_2 \leq 1.$$

Proof of Lemma C.7. Let $\boldsymbol{y} \in \mathbb{C}^N$. Then

$$\begin{aligned}
\left\|\overline{\boldsymbol{\Psi}}^H \boldsymbol{y}\right\|_2^2 &= \sum_{i=0}^{J-1} \sum_{l_i=0}^{k_i-1} \left| \left\langle \boldsymbol{y}, \boldsymbol{E}_{f_i} \sqrt{\lambda_{N,W_i}^{(\ell_i)}} \boldsymbol{s}_{N,W_i}^{(\ell_i)} \right\rangle \right|^2 \\
&= \sum_{i=0}^{J-1} \sum_{l_i=0}^{k_i-1} \left\langle \boldsymbol{y}, \boldsymbol{E}_{f_i} \sqrt{\lambda_{N,W_i}^{(\ell_i)}} \boldsymbol{s}_{N,W_i}^{(\ell_i)} \right\rangle \left\langle \boldsymbol{E}_{f_i} \sqrt{\lambda_{N,W_i}^{(\ell_i)}} \boldsymbol{s}_{N,W_i}^{(\ell_i)}, \boldsymbol{y} \right\rangle \\
&= \sum_{i=0}^{J-1} \sum_{l_i=0}^{k_i-1} \boldsymbol{y}^H \boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)} \lambda_{N,W_i}^{(\ell_i)} (\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \boldsymbol{E}_{f_i}^H \boldsymbol{y} \\
&\leq \sum_{i=0}^{J-1} \sum_{l_i=0}^{N-1} \boldsymbol{y}^H \boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)} \lambda_{N,W_i}^{(\ell_i)} (\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \boldsymbol{E}_{f_i}^H \boldsymbol{y} \\
&= \sum_{i=0}^{J-1} \boldsymbol{y}^H \boldsymbol{E}_{f_\ell} \mathcal{I}_N(\mathcal{B}_{W_i}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}^H \boldsymbol{y}))) = \sum_{i=0}^{J-1} \langle \mathcal{I}_N(\mathcal{B}_{W_i}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}^H \boldsymbol{y}))), \boldsymbol{E}_{f_i}^H \boldsymbol{y} \rangle \\
&= \sum_{i=0}^{J-1} \langle \mathcal{B}_{W_i}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}^H \boldsymbol{y})), \mathcal{I}_N^*(\boldsymbol{E}_{f_i}^H \boldsymbol{y}) \rangle = \sum_{i=0}^{J-1} \left\| \mathcal{B}_{W_i}(\mathcal{I}_N^*(\boldsymbol{E}_{f_i}^H \boldsymbol{y})) \right\|_2^2 \\
&= \sum_{i=0}^{J-1} \int_{f_i - W_i}^{f_i + W_i} |\widetilde{\boldsymbol{y}}(f)|^2 df = \int_{-1/2}^{1/2} \left(\sum_{i=0}^{J-1} \mathbb{1}_{[f_i - W_i, f_i + W_i)}(f) \right) |\widetilde{\boldsymbol{y}}(f)|^2 df,
\end{aligned}$$

where the fourth line follows because

$$\boldsymbol{y}^H \boldsymbol{E}_{f_i} \boldsymbol{s}_{N,W_i}^{(\ell_i)} \lambda_{N,W_i}^{(\ell_i)} (\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \boldsymbol{E}_{f_i}^H \boldsymbol{y} = \left\| \sqrt{\lambda_{N,W_i}^{(\ell_i)}} (\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \boldsymbol{E}_{f_i}^H \boldsymbol{y} \right\|_2^2 \geq 0,$$

the fifth line follows because $\sum_{\ell_i=0}^{N-1} \boldsymbol{s}_{N,W_i}^{(\ell_i)} \lambda_{N,W_i}^{(\ell_i)} (\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H \boldsymbol{x} = \mathcal{I}_N(\mathcal{B}_{W_i}(\mathcal{I}_N^*(\boldsymbol{x})))$, and we use $\widetilde{\boldsymbol{y}}(f) = \sum_{n=0}^{N-1} \boldsymbol{y}[n] e^{-j2\pi f n}$ as the DTFT of $\mathcal{I}_N^*(\boldsymbol{y})$ in the last three equations.

Noting that $\sum_{i=0}^{J-1} \mathbb{1}_{[f_i - W_i, W_i + f_i)}(f) \leq 1$ for all $f \in [-\frac{1}{2}, \frac{1}{2}]$ since we assume there is no overlap between each interval $[f_i - W_i, W_i + f_i)$, we conclude

$$\left\|\overline{\boldsymbol{\Psi}}^H \boldsymbol{y}\right\|_2^2 \leq \int_{-1/2}^{1/2} |\widetilde{\boldsymbol{y}}(f)|^2 df = \|\boldsymbol{y}\|_2^2$$

and

$$||\overline{\boldsymbol{\Psi}}||_2 \leq 1.$$

□

Lemma C.8. *For any $k_i \in \{1, 2, \ldots, N\}, i \in [J]$, let $\boldsymbol{\Psi}$ and $\overline{\boldsymbol{\Psi}}$ be the matrices defined in (5.1) and (C.6) respectively. Then for any $\boldsymbol{y} \in \mathbb{C}^{N\times 1}$,*

$$||\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{y}||_2 \geq ||\overline{\boldsymbol{\Psi}}\ \overline{\boldsymbol{\Psi}}^H\boldsymbol{y}||_2. \tag{C.7}$$

Proof of Lemma C.8. Let $\overline{\boldsymbol{\Psi}} = \boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}\Sigma_{\overline{\boldsymbol{\Psi}}}\boldsymbol{V}_{\overline{\boldsymbol{\Psi}}}^H$ be a reduced SVD of $\overline{\boldsymbol{\Psi}}$, where both $\boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}$ and $\boldsymbol{V}_{\overline{\boldsymbol{\Psi}}}$ are orthonormal matrices of the proper dimension, and $\boldsymbol{\Sigma}_{\overline{\boldsymbol{\Psi}}}$ is a diagonal matrix whose diagonal elements are the non-zero singular values of $\overline{\boldsymbol{\Psi}}$. We have

$$\begin{aligned}
||\overline{\boldsymbol{\Psi}}\ \overline{\boldsymbol{\Psi}}^H\boldsymbol{y}||_2 &= ||\boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}\Sigma_{\overline{\boldsymbol{\Psi}}}^2\boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}^H\boldsymbol{y}||_2 \\
&\leq ||\boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}^H\boldsymbol{y}||_2 \\
&= ||\boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}\boldsymbol{U}_{\overline{\boldsymbol{\Psi}}}^H\boldsymbol{y}||_2 \\
&= ||\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{y}||_2
\end{aligned}$$

where the second lines follows because $||\overline{\boldsymbol{\Psi}}||_2 \leq 1$ and hence the diagonal elements $\boldsymbol{\Sigma}_{\overline{\boldsymbol{\Psi}}}$ are bounded above by 1, and the fourth line follows because each column in $\overline{\boldsymbol{\Psi}}$ is in also $\boldsymbol{\Psi}$ and hence $||\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{y}||_2 = ||\boldsymbol{P}_{\boldsymbol{U}_{\overline{\Psi}}}\boldsymbol{y}||_2$. □

Now we turn to prove Theorem 5.5. By (C.7), we observe that

$$\begin{aligned}
&\left\|\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 \geq \left\|\overline{\boldsymbol{\Psi}}\ \overline{\boldsymbol{\Psi}}^H\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 = \left\|\sum_{i=0}^{J-1}\sum_{\ell_i=0}^{k_i-1}\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)}\lambda_{N,W_i}^{(\ell_i)}(\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H\boldsymbol{E}_{f_i}^H\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 \\
&= \left\|\boldsymbol{B}_{N,\mathbb{W}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)} - \sum_{i=0}^{J-1}\sum_{\ell_i=k_i}^{N-1}\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)}\lambda_{N,W_i}^{(\ell_i)}(\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H\boldsymbol{E}_{f_i}^H\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 \\
&\geq \left\|\boldsymbol{B}_{N,\mathbb{W}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 - \sum_{i=0}^{J-1}\sum_{\ell_i=k_i}^{N-1}\left\|\boldsymbol{E}_{f_i}\boldsymbol{s}_{N,W_i}^{(\ell_i)}\lambda_{N,W_i}^{(\ell_i)}(\boldsymbol{s}_{N,W_i}^{(\ell_i)})^H\boldsymbol{E}_{f_i}^H\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 \geq \lambda_{N,\mathbb{W}}^{(\ell)} - \sum_{i=0}^{J-1}\sum_{l_i=k_i}^{N-1}\lambda_{N,W_i}^{(\ell_i)}.
\end{aligned}$$

C.7 Proof of Corollary 5.1

It follows from Theorem 5.5 that

$$\begin{aligned}\left\|\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}\right\|_2 &\geq \lambda_{N,\mathbb{W}}^{(\ell)} - \sum_{i=0}^{J-1}\sum_{l_i=k_i}^{N-1}\lambda_{N,W_i}^{(\ell_i)} \\ &\geq 1 - \overline{C}_1(\mathbb{W},\epsilon)N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N} - \sum_{i=0}^{J-1}\sum_{l_i=k_i}^{N-1} C_3(W_i,\epsilon)e^{-C_4(W_i,\epsilon)N} \\ &\geq 1 - \overline{C}_1(\mathbb{W},\epsilon)N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N} - \sum_{i=0}^{J-1}\sum_{l_i=k_i}^{N-1} \frac{1}{J}\overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N} \\ &\geq 1 - \overline{C}_1(\mathbb{W},\epsilon)N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N} - N\overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}\end{aligned}$$

for all $N \geq \max\{\overline{N}_0(\mathbb{W},\epsilon), \overline{N}_1(\mathbb{W},\epsilon)\}$, where the second line follows by bounding the $\lambda_{N,\mathbb{W}}^{(\ell)}$ term using Theorem 5.3 and by bounding the $\lambda_{N,W_i}^{(\ell_i)}$ terms using Lemma 2.1, and the third line follows because $\overline{C}_3(\mathbb{W},\epsilon) = J\max\{C_3(W_i,\epsilon),\ \forall\ i \in [J]\}$ and $\overline{C}_4(\mathbb{W},\epsilon) = \min\{C_4(W_i,\epsilon),\ \forall\ i \in [J]\}$.

Let $\kappa_2(N,\mathbb{W},\epsilon) = \overline{C}_1(\mathbb{W},\epsilon)N^2 e^{-\overline{C}_2(\mathbb{W},\epsilon)N} + N\overline{C}_3(\mathbb{W},\epsilon)e^{-\overline{C}_4(\mathbb{W},\epsilon)N}$. Then $||\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)} - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}||_2^2 \leq 2\kappa_2(N,\mathbb{W},\epsilon) - \kappa_2^2(N,\mathbb{W},\epsilon)$. Noting also that $\langle \boldsymbol{u}_{N,\mathbb{W}}^{(\ell)}, \boldsymbol{u}_{N,\mathbb{W}}^{(k)}\rangle = 0$ for all $k \neq l$, (5.5) follows directly from Lemma C.6.

C.8 DTFT of DPSS vectors

The results presented in this appendix are useful in Appendix C.9, where we analyze the performance of the DPSS vectors for representing sampled pure tones inside the band of interest. Let $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)$ denote the DTFT of the sequence $\mathcal{T}_N(s_{N,W}^{(\ell)})$, i.e., $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f) = \sum_{n=0}^{N-1} s_{N,W}^{(\ell)}[n]e^{-j2\pi fn}$. Figure C.1 shows $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)$ for all $\ell \in [N]$ with $N = 1024$ and $W = \frac{1}{4}$. We observe that the first $\approx 2NW$ DPSS vectors have their spectrum mostly concentrated in $[-W, W]$, only a small fraction of DPSS vectors whose indices are near $2NW$ have a relatively flat spectrum over $[-\frac{1}{2}, \frac{1}{2}]$, and the remaining DPSS vectors have their spectrum mostly concentrated outside of the band $[-W, W]$. This phenomenon is captured formally in the asymptotic expressions for $\lambda_{N,W}^{(\ell)}$ and $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)$ from [118].

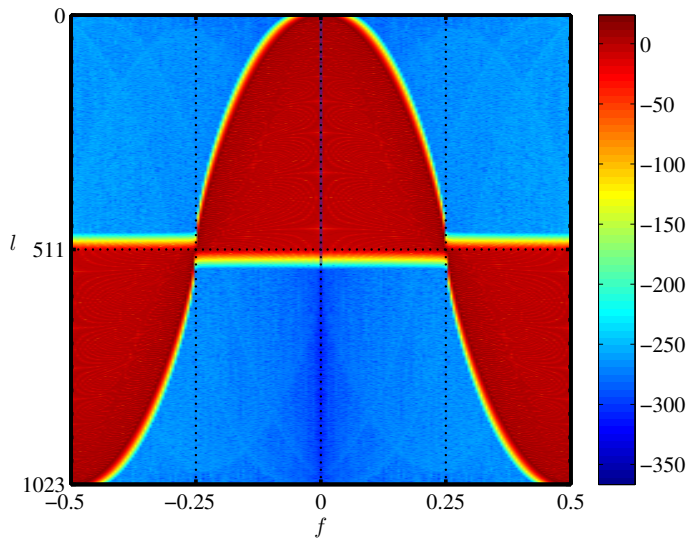


Figure C.1: Illustration of $\left|\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)\right|^2$, or the energy in $\{\boldsymbol{e}_f\}$ captured by each DPSS vector. The horizontal axis stands for the digital frequency f, which ranges over $[-\frac{1}{2}, \frac{1}{2}]$, while the vertical axis stands for the index $\ell \in [N]$. The ℓ-th horizontal line shows $10\log_{10}\left|\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)\right|^2$. Here $N = 1024$ and $W = \frac{1}{4}$.

Lemma C.9. *([118]) Fix $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, 1)$. Let $\alpha := 1 - A = 1 - \cos 2\pi W$.*

1. *For fixed ℓ, as $N \to \infty$, we have*

$$1 - \lambda_{N,W}^{(\ell)} \sim c_5^2 / \left(2\sqrt{2\alpha}\right)$$

and

$$\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f) \sim \begin{cases} c_3 f_4(f), & W \le |f| \le \arccos(A - N^{-3/2})/2\pi, \\ c_5 f_5(f), & \arccos(A - N^{-3/2})/2\pi \le |f| \le 1/2. \end{cases}$$

Here

$$
\begin{aligned}
c_5 &= (\ell!)^{-1/2}\pi^{1/4}2^{(14\ell+15)/8}\alpha^{(2l+3)/8}N^{(2\ell+1)/4}(\sqrt{2}+\sqrt{\alpha})^{-N}(2-\alpha)^{(N-\ell-1/2)/2} \\
&= (\ell!)^{-1/2}\pi^{1/4}2^{(14\ell+15)/8}\alpha^{(2l+3)/8}N^{(2\ell+1)/4}(2-\alpha)^{-(\ell+1/2)/2}e^{-\frac{\gamma}{2}N}, \\
c_3 &= \pi^{1/2}2^{-1/2}\alpha^{-1/4}[2-\alpha]^{-1/4}N^{1/2}c_5 = O(N^{1/2})c_5, \\
\gamma &= \log(1+\frac{2\sqrt{\alpha}}{\sqrt{2}-\sqrt{\alpha}}), \\
f_4(f) &= J_0\left(\frac{N}{\sqrt{2-\alpha}}\sqrt{A-\cos(2\pi f)}\right), \\
f_5(f) &= \frac{\cos\left(\frac{N}{2}\arcsin(\theta(f))+\frac{1}{2}(l+\frac{1}{2})\arcsin(\phi(f))+(l-N)\frac{\pi}{4}+\frac{3\pi}{8}\right)}{((A-\cos(2\pi f))(1-\cos(2\pi f)))^{1/4}}, \\
\theta(f) &= \frac{\alpha+2\cos(2\pi f)}{2-\alpha},\ \phi(f) = \frac{(2-3\alpha)-(2+\alpha)\cos(2\pi f)}{(2-\alpha)(1-\cos(2\pi f))},
\end{aligned}
$$

where J_0 is the Bessel function of the first kind.

2. *As $N\to\infty$ and with $\ell = \lfloor 2NW(1-\epsilon')\rfloor$ for any $\epsilon'\in(0,\epsilon]$, we have*

$$1-\lambda^{(\ell)}_{N,W} \sim 2\pi L_2^{-1}d_6^2$$

and

$$\widetilde{\boldsymbol{s}}^{(\ell)}_{N,W}(f) \sim \begin{cases} d_4 g_5(f), & W\le|f|\le \arccos(A-N^{-1})/2\pi, \\ d_6 g_6(f), & \arccos(A-N^{-1})/2\pi\le|f|\le 1/2. \end{cases}$$

Here

$$
\begin{aligned}
d_6 &= (L_2)^{-1/2}\pi^{1/2}2^{1/2}e^{-CL_4/4}e^{-NL_3/2}, \\
d_4 &= (L_2)^{-1/2}\pi(1-A^2)^{-1/4}e^{-CL_4/4}e^{-NL_3/2}N^{1/2}, \\
g_5(f) &= J_0\left(N\sqrt{\frac{B-A}{1-A^2}(\cos(2\pi f)-A)}\right), \\
g_6(f) &= R(f)\cos\left(\pi N\int_f^{1/2}\sqrt{\frac{B-\cos(2\pi t)}{A-\cos(2\pi t)}}dt\right. \\
&\qquad \left. +\frac{\pi C}{2}\int_f^{1/2}\frac{dt}{\sqrt{(B-\cos(2\pi t))(A-\cos(2\pi t))}}+\theta\right),
\end{aligned}
$$

$$
\begin{aligned}
R(f) &= \left|\left(B-\cos(2\pi f)\right)\left(A-\cos(2\pi f)\right)\right|^{-1/4}, \\
C &= \frac{1}{L_2} \mod \left(\frac{N}{2}L_1 + \left(2+(-1)^l\right)\frac{\pi}{4}, 2\pi\right), \\
\theta &= \mod \left(\frac{\pi}{4} - \frac{N}{2}L_5 - \frac{C}{4}L_6, 2\pi\right), \\
L_1 &= \int_B^1 P(\xi)d\xi,\ L_2 = \int_B^1 Q(\xi)d\xi,\ L_3 = \int_A^B P(\xi)d\xi,\ L_4 = \int_A^B Q(\xi)d\xi \\
L_5 &= \int_{-1}^A P(\xi)d\xi,\ L_6 = L_2, \\
P(\xi) &= \left|\frac{\xi-B}{(\xi-A)\left(1-\xi^2\right)}\right|^{1/2},\ Q(\xi) = \left|\left(\xi-B\right)\left(\xi-A\right)\left(1-\xi^2\right)\right|^{-1/2},
\end{aligned}
$$

where B is determined so that $\int_B^1 \sqrt{\frac{\xi-B}{(\xi-A)(1-\xi^2)}}d\xi = \frac{l}{N}\pi$ and $\mod(y, 2\pi)$ returns the remainder after division of y by 2π.

C.9 Proof of Theorem 5.6

Noting that $\boldsymbol{S}_{N,W}$ forms an orthobasis for $\mathbb{C}^{N\times N}$, the main idea is to show that the DPSS vectors $\boldsymbol{s}_{N,W}^{(2NW(1+\epsilon))}, \boldsymbol{s}_{N,W}^{(2NW(1+\epsilon)+1)}, \ldots, \boldsymbol{s}_{N,W}^{(N-1)}$ have their spectrum most concentrated outside of the band $[-W, W]$.

Since the sequence $s_{N,W}^{(\ell)}$ is exactly bandlimited to the frequency range $|f| \le W$, we know that its DTFT $\widetilde{s}_{N,W}^{(\ell)}(f) := \sum_{n=-\infty}^{\infty} s_{N,W}^{(\ell)}[n]e^{j2\pi fn}$ vanishes for all $W < |f| < \frac{1}{2}$. By noting that the first $\approx 2NW$ DPSS's are also approximately time-limited to the index range $n = 0, 1, \ldots, N-1$, we may expect that $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f) := \sum_{n=0}^{N-1} \boldsymbol{s}_{N,W}^{(\ell)}[n]e^{j2\pi fn}$ is also approximately 0 for all $W < |f| < \frac{1}{2}$ and $\ell \le 2NW(1-\epsilon)$. This illustrates informally why the DTFT of the first $\approx 2NW$ DPSS vectors is concentrated inside the band $[-W, W]$. By employing the antisymmetric property [118] which states that $|\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)| = |\widetilde{\boldsymbol{s}}_{N,\frac{1}{2}-W}^{(N-1-\ell)}(\frac{1}{2}-f)|$, we then have that the DPSS vectors $\boldsymbol{s}_{N,W}^{(2NW(1+\epsilon))}$, $\boldsymbol{s}_{N,W}^{(2NW(1+\epsilon)+1)}$, $\ldots$, $\boldsymbol{s}_{N,W}^{(N-1)}$ are almost orthogonal to any sinusoid with frequency inside the band $[-W, W]$.

Recall that $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)$ is the DTFT of the sequence $\mathcal{T}_N(s_{N,W}^{(\ell)})$. We have

$$\langle \boldsymbol{s}_{N,W}^{(\ell)}, \boldsymbol{e}_f \rangle = \widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f),$$

for all $\ell \in [N]$. As we have observed in Figure C.1, the spectrum of the first $\approx 2NW$ DPSS vectors is approximately concentrated on the frequency interval $[-W, W]$. This behavior is captured formally in the following results.

Corollary C.1. *Let $A = \cos 2\pi W$. For fixed $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \min(\frac{1}{2W} - 1, 1))$, there exists a constant $C_6(W, \epsilon)$ (which may depend on W and ϵ) such that*

$$|\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)| \leq C_6(W,\epsilon) N^{3/4} e^{-\frac{C_2(W,\epsilon)}{2} N}, \quad W \leq |f| \leq 1/2$$

for all $N \geq N_0(W, \epsilon)$ and $\ell \leq 2NW(1-\epsilon)$. Here $C_2(W, \epsilon)$ and $N_0(N, \epsilon)$ are constants specified in Lemma 2.1.

Proof of Corollary C.1. The main approach is to bound $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f),\ W \leq |f| \leq 1/2$ with the expressions presented in Lemma C.9. Suppose $\epsilon \in (0, 1)$ is fixed.

- For fixed ℓ and large N:

  In order to quantify the decay rate of $|\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)|$, we exploit some results concerning of $f_4(f)$ from [100] and $f_5(f)$ as follows:

  $$|J_0(x)| \leq 1,\ \forall\ x \geq 0, \tag{C.8}$$

  and for any $\frac{\arccos(A - N^{-3/2})}{2\pi} \leq |f| \leq 1/2$, one may verify that

$$\begin{aligned} |f_5(f)| &\leq \frac{1}{\left((A - \cos(2\pi f))(1 - \cos(2\pi f))\right)^{1/4}} \\ &\leq \frac{1}{\left((A - (A - N^{-3/2}))\left(1 - (A - N^{-3/2})\right)\right)^{1/4}} \\ &\leq \frac{1}{\left((N^{-3/2}))(N^{-3/2}))\right)^{1/4}} = N^{3/4}, \end{aligned}$$

where the last line follows because $1 - A \geq 0$.

Recall that $c_3 = \pi^{1/2} 2^{-1/2} \alpha^{-1/4} (2-\alpha)^{-1/4} N^{1/2} c_5$ and $c_5 \sim \sqrt{2\sqrt{2\alpha}\left(1 - \lambda^{(\ell)}_{N,W}\right)}$. Plugging these into Lemma C.9 and utilizing Lemma 2.1, we get the exponential decay of $|\widetilde{\boldsymbol{s}}^{(\ell)}_{N,W}(f)|$, $|f| \geq W$ as

$$|\widetilde{\boldsymbol{s}}^{(\ell)}_{N,W}(f)| \leq \begin{cases} C_7'(W,\epsilon) N^{1/2} e^{-\frac{C_2}{2}N}, & W \leq |f| \leq \arccos\left(A - N^{-3/2}\right)/2\pi, \\ C_8'(W,\epsilon) N^{3/4} e^{-\frac{C_2}{2}N}, & \arccos\left(A - N^{-3/2}\right)/2\pi \leq |f| \leq 1/2, \end{cases}$$

for fixed ℓ and $N \geq N_0(W,\epsilon)$. Here $C_7'(W,\epsilon) = \pi^{1/2} 2^{1/4} (2-\alpha)^{-1/4} \sqrt{C_1(W,\epsilon)}$, $C_8'(W,\epsilon) = (2\sqrt{2\alpha} C_1(W,\epsilon))^{1/2}$, and $N_0(W,\epsilon)$, $C_1(W,\epsilon)$ and $C_2(W,\epsilon)$ are constants as specified in Lemma 2.1.

- For large N and $\ell = \lfloor 2NW(1-\epsilon') \rfloor$, $\forall\ \epsilon' \in (0,\epsilon]$:

  Note that $\int_B^1 \sqrt{\frac{\xi - B}{(\xi - A)(1-\xi^2)}} d\xi$ is a decreasing function of B and $\int_A^1 \sqrt{\frac{\xi - A}{(\xi - A)(1-\xi^2)}} d\xi = 2W\pi > \frac{l}{N}\pi$. Hence $1 > B > A$. Now we have

$$|g_6(f)| \leq |R(f)| \leq \frac{1}{(A - \cos(2\pi f))^{1/2}} \leq \frac{1}{(A - (A - N^{-1}))^{1/2}} \leq N^{1/2}$$

  for all $\arccos(A - N^{-1})/2\pi \leq |f| \leq 1/2$.

  Recall that $|g_5(f)| \leq= 1$ from (C.8), $d_4 = \pi^{1/2}(1-A^2)^{-1/4} 2^{-1/2} N^{1/2} d_6$ and $d_6 \sim \sqrt{\frac{1 - \lambda^{(\ell)}_{N,W}}{2\pi}}$. Plugging these into Lemma C.9 and utilizing the bound on $\lambda^{(\ell)}_{N,W}$ in Lemma 2.1, we get the exponential decay of $|\widetilde{\boldsymbol{s}}^{(\ell)}_{N,W}(f)|$, $|f| \geq W$ as

$$|\widetilde{\boldsymbol{s}}^{(\ell)}_{N,W}(f)| \leq \begin{cases} C_7''(W,\epsilon) N^{1/2} e^{-\frac{C_2}{2}N}, & W \leq |f| \leq \arccos[A - N^{-1}]/2\pi, \\ C_8''(W,\epsilon) N^{1/2} e^{-\frac{C_2}{2}N}, & \arccos[A - N^{-1}]/2\pi \leq |f| \leq 1/2, \end{cases}$$

  for all $\ell = \lfloor 2NW(1-\epsilon') \rfloor$, $\forall\ \epsilon' \in (0,\epsilon]$ and $N \geq N_0(W,\epsilon)$. Here $C_8''(W,\epsilon) = \sqrt{C_1(W,\epsilon)/2\pi}$, $C_7''(W,\epsilon) = 2^{-1}(1-A^2)^{-1/4}\sqrt{C_1(W,\epsilon)}$, and $N_0(W,\epsilon)$, $C_1(W,\epsilon)$ and $C_2(W,\epsilon)$ are constants as specified in Lemma 2.1.

Set

$$C_6(W,\epsilon) = \max\{C_7'(W,\epsilon), C_8'(W,\epsilon), C_7''(W,\epsilon), C_8''(W,\epsilon)\}$$
$$= \max\left\{\pi^{1/2}\left(\frac{2}{2-\alpha}\right)^{1/4}, 2^{-1}(1-A^2)^{-1/4}\right\}\sqrt{C_1(W,\epsilon)}.$$

This completes the proof of Corollary C.1. □

Lemma C.10. *([118]) For fixed $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2W} - 1)$, $\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)$ and $\widetilde{\boldsymbol{s}}_{N,\frac{1}{2}-W}^{(N-1-\ell)}(f)$ satisfy*

$$|\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)| = |\widetilde{\boldsymbol{s}}_{N,\frac{1}{2}-W}^{(N-1-\ell)}(\frac{1}{2} - f)|$$

for all $\ell \geq 2NW(1+\epsilon)$.

Now we can conclude that $\langle \boldsymbol{e}_f, \boldsymbol{s}_{N,W}^{(\ell)}\rangle$ decays exponentially in N for all $\ell \geq 2NW(1+\epsilon)$ and $|f| \leq W$ by combining the above results.

Corollary C.2. *Fix $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2W} - 1)$. Let $W' = \frac{1}{2} - W$ and $\epsilon' = \frac{W}{\frac{1}{2}-W}\epsilon$. Then*

$$|\langle \boldsymbol{e}_f, \boldsymbol{s}_{N,W}^{(\ell)}\rangle| = |\widetilde{\boldsymbol{s}}_{N,W}^{(\ell)}(f)| \leq C_6(W',\epsilon')N^{3/4}e^{-\frac{C_2(W',\epsilon')}{2}N}, \ \forall |f| \leq W$$

for all $N \geq N_0(W',\epsilon')$ and all $\ell \geq 2NW(1+\epsilon)$. Here, $C_2(W',\epsilon')$ and $N_0(W',\epsilon')$ are constants specified in Lemma 2.1 with respect to W' and ϵ', and $C_6(W',\epsilon')$ is the constant specified in Corollary C.1 with respect to W' and ϵ'.

Proof of Corollary C.2. Let $\ell' = N - 1 - \ell$. For all $\ell \geq 2NW(1+\epsilon)$, we have

$$\ell' = N - 1 - \ell \leq N - 2NW(1+\epsilon) = 2N(\frac{1}{2} - W)(1 - \frac{W}{\frac{1}{2}-W}\epsilon).$$

Let $W' = \frac{1}{2} - W$ and $\epsilon' = \frac{W}{\frac{1}{2}-W}\epsilon \in (0,1)$. It follows from from Corollary C.1 and Lemma C.10 that

$$|\langle \boldsymbol{e}_f, \boldsymbol{s}_{N,W}^{(\ell)}\rangle| = |\langle \boldsymbol{e}_{\frac{1}{2}-f}, \boldsymbol{s}_{N,W'}^{(\ell')}\rangle| \leq C_6(W',\epsilon')N^{3/4}e^{-\frac{C_2(W',\epsilon')}{2}N}, \ \forall\ |f| \leq W$$

for all $N \geq N_0(W', \epsilon')$. $\square$

Recall that $C_6(W', \epsilon') = \max\left\{\pi^{1/2}\left(\frac{2}{\alpha}\right)^{1/4}, 2^{-1}(1-A^2)^{-1/4}\right\}\sqrt{C_1(W', \epsilon')}$, $A = \cos(2\pi W)$ and $\alpha = 1 - A$. As W gets closer to 0 or $\frac{1}{2}$, the variable $(1-A^2)^{-1/4}$ becomes larger, and we have $(1-A^2)^{-1/4} \to 1/\sqrt{2\pi W}$ as $W \to 0$. Also we have $\left(\frac{2}{\alpha}\right)^{1/4} \to 1/\sqrt{\pi W}$ as $W \to 0$. Therefore, for any non-negligible bandwidth which is the main assumption in this paper, the variable $\max\left\{\pi^{1/2}\left(\frac{2}{\alpha}\right)^{1/4}, 2^{-1}(1-A^2)^{-1/4}\right\}\sqrt{C_1(W', \epsilon')}$ would not be too large.

Now, for fixed $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2W} - 1)$, we have

$$\begin{aligned}
||\boldsymbol{e}_f - \boldsymbol{P}_{[\boldsymbol{S}_{N,W}]_k}\boldsymbol{e}_f||_2^2 &= \sum_{l=2NW(1+\epsilon)}^{N-1} |\langle \boldsymbol{e}_f, \boldsymbol{s}_{N,W}^{(\ell)}\rangle|^2 \\
&\leq \sum_{l=2NW(1+\epsilon)}^{N-1} C_6^2(W', \epsilon') N^{3/2} e^{-C_2(W', \epsilon')N} \\
&\leq C_9(W', \epsilon') N^{5/2} e^{-C_2(W', \epsilon')N}
\end{aligned}$$

for all $|f| \leq W$ and $N \geq N_0(W', \epsilon')$, where $C_9(W', \epsilon') = C_6^2(W', \epsilon')$.

C.10 Proof of Corollary 5.2

Suppose $f \in [f_i - W_i, f_i + W_i]$ for some particular $i \in [J]$. Denote by

$$C_{10}(\mathbb{W}, \epsilon) = \max\{C_9(W_i', \epsilon'), \forall i \in [J]\}, C_{11}(\mathbb{W}, \epsilon) = \min\{C_2(W_i', \epsilon'), \forall i \in [J]\}.$$

It follows from Theorem 5.6 that

$$\begin{aligned}
||\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f||_2^2 &\leq ||\boldsymbol{e}_f - \boldsymbol{P}_{[\boldsymbol{E}_{f_i}\boldsymbol{S}_{N,W_i}]_{2NW_i(1+\epsilon)}}\boldsymbol{e}_f||_2^2 \\
&= ||\boldsymbol{e}_{f-f_i} - \boldsymbol{P}_{[\boldsymbol{S}_{N,W_i}]_{2NW_i(1+\epsilon)}}\boldsymbol{e}_{f-f_i}||_2^2 \\
&\leq C_9(W_i', \epsilon') N^{5/2} e^{-C_2(W_i', \epsilon')N} \leq C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon)N}
\end{aligned}$$

for all $N \geq N_0(W_i', \epsilon')$. We complete the proof by setting $N_2(\mathbb{W}, \epsilon) = \max\{N_0(W_i', \epsilon'), \forall i \in [J]\}$.

C.11 Proof of Theorem 5.7

We first present the following useful result from [32].

Theorem C.1. *([32] Theorem 4.2) Suppose x is a continuous, zero-mean, wide sense stationary random process with power spectrum*

$$P_x(F) = \begin{cases} \frac{1}{B_{\text{band}}}, & F \in [F_c - \frac{B_{\text{band}}}{2}, F_c + \frac{B_{\text{band}}}{2}], \\ 0, & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{x} = [x(0)\ x(T_s)\ \dots\ x((N-1)T_s)]^T \in \mathbb{C}^N$ denote a finite vector of samples acquired from $x(t)$ with a sampling interval of $T_s \le 1/(2\max\{|F_c \pm \frac{B_{\text{band}}}{2}|\})$. Let $f_c = F_c T_s$ and $W = \frac{B_{\text{band}} T_s}{2}$. We will have

$$\mathbb{E}\left[||\boldsymbol{x} - \boldsymbol{P_Q x}||_2^2\right] = \frac{1}{2W}\int_{f_c-W}^{f_c+W} ||\boldsymbol{e}_f - \boldsymbol{P_Q e}_f||_2^2 df = \frac{1}{2W}\sum_{l=k}^{N-1} \lambda_{N,W}^{(l)}.$$

Furthermore, for fixed $\epsilon \in (0, \frac{1}{2W} - 1)$, set $k = 2NW(1+\epsilon)$. Then

$$\mathbb{E}\left[||\boldsymbol{x} - \boldsymbol{P_Q x}||_2^2\right] \le \frac{C_3(W,\epsilon)}{2W} N e^{-C_4(W,\epsilon)N} \tag{C.9}$$

for all $N \ge N_1(W,\epsilon)$, where $N_1(W,\epsilon)$, $C_3(W,\epsilon)$, $C_4(W,\epsilon)$ are constants specified in Lemma 2.1. For comparison, $\mathbb{E}\left[||\boldsymbol{x}||_2^2\right] = ||\boldsymbol{e}_f||_2^2 = N$.

Since $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_{J-1}$ are independent and zero-mean, we have

$$\begin{aligned} \mathbb{E}\left[\|\boldsymbol{x}\|_2^2\right] &= \sum_{n=0}^{N-1} \mathbb{E}\left[|\boldsymbol{x}[n]|^2\right] = \sum_{n=0}^{N-1} \sum_{0\le i,i'\le J-1} \mathbb{E}\left[\boldsymbol{x}_i[n]\overline{\boldsymbol{x}_i'[n]}\right] \\ &= \sum_{n=0}^{N-1}\sum_{i=0}^{J-1} \mathbb{E}\left[|\boldsymbol{x}_i[n]|^2\right] = N\sum_{i=0}^{J-1}\frac{1}{J} = N. \end{aligned}$$

Applying Theorem C.1, we acquire

$$\mathbb{E}\left[\left\|\boldsymbol{x}_i - \boldsymbol{P}_{\left[\boldsymbol{E}_{f_i}\boldsymbol{S}_{N,W_i}\right]_{k_i}}\boldsymbol{x}\right\|_2^2\right] = \frac{1}{|\mathbb{W}|}\sum_{l=k_i}^{N-1}\lambda_{N,W_i}^{(\ell)}.$$

Note that the power spectrum $P_{x_i}(F)$ assumed in (5.7) results in the constant $\frac{1}{|\mathbb{W}|}$ instead of $\frac{1}{2W_i}$.

Now, we have

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{x}-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}\|_2^2\right] &= \mathbb{E}\left[\left\|\sum_{i=0}^{J-1}\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}(\sum_{i=0}^{J-1}\boldsymbol{x}_i)\right\|_2^2\right] = \mathbb{E}\left[\left\|\sum_{i=0}^{J-1}(\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i)\right\|_2^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=0}^{J-1}(\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i)^H\right)\left(\sum_{i=0}^{J-1}(\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i)\right)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{J-1}\|\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i\|_2^2+\sum_{i=0}^{J-1}\sum_{i'=0,i'\neq i}^{J-1}(\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i)^H(\boldsymbol{x}_{i'}-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_{i'})\right] \\
&= \sum_{i=0}^{J-1}\mathbb{E}\left[\|\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i\|_2^2\right]+\sum_{i=0}^{J-1}\sum_{i'=0,i'\neq i}^{J-1}\mathbb{E}\left[(\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i)^H(\boldsymbol{x}_{i'}-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_{i'})\right] \\
&= \sum_{i=0}^{J-1}\mathbb{E}\left[\|\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i\|_2^2\right]+\sum_{i=0}^{J-1}\sum_{i'=0,i'\neq i}^{J}\mathbb{E}\left[\boldsymbol{x}_i^H\boldsymbol{x}_{i'}-\boldsymbol{x}_i^H\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_{i'}\right] \\
&= \sum_{i=0}^{J-1}\mathbb{E}\left[\|\boldsymbol{x}_i-\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i\|_2^2\right] \leq \sum_{i=0}^{J-1}\mathbb{E}\left[\left\|\boldsymbol{x}_i-\boldsymbol{P}_{[\boldsymbol{E}_{f_i}\boldsymbol{S}_{N,W_i}]_{k_i}}\boldsymbol{x}_i\right\|_2^2\right] \\
&= \sum_{i=0}^{J-1}\frac{1}{|\mathbb{W}|}\sum_{l=k_i}^{N-1}\lambda_{N,W_i}^{(\ell)}
\end{aligned}$$

where the equality in the sixth line follows because $\mathbb{E}\left[\boldsymbol{x}_{i'}^H\boldsymbol{x}_i\right] = \left(\mathbb{E}\left[\boldsymbol{x}_{i'}\right]\right)^H\left(\mathbb{E}\left[\boldsymbol{x}_i\right]\right) = 0$ and $\mathbb{E}\left[\boldsymbol{x}_{i'}^H\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i\right] = \left(\mathbb{E}\left[\boldsymbol{x}_{i'}\right]\right)^H\left(\mathbb{E}\left[\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}_i\right]\right) = 0$ for all $i', i \in [J], i' \neq i$, and the inequality in the sixth line follows because the column space of $[\boldsymbol{E}_{f_i}\boldsymbol{S}_{N,W_i}]_{k_i}$ is inside the column space of $\boldsymbol{\Psi}$ for all $i \in [J]$.

C.12 Proof of Corollary 5.3

It is useful to express the sampled bandpass signal $\boldsymbol{x}$ as

$$\boldsymbol{x} = \int_{\mathbb{W}} \widetilde{x}(f)\boldsymbol{e}_f \,\mathrm{d}\, f, \tag{C.10}$$

where we recall that $\widetilde{x}(f)$ denotes the DTFT of $x[n]$, which is the infinite-length sequence that one obtains by uniformly sampling $x(t)$ with sampling rate T_s.

Now it follows from (C.10) that

$$\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{x}\|_2^2 &= \left\| \int_{\mathbb{W}} \widetilde{x}(f)\boldsymbol{e}_f df - \int_{\mathbb{W}} \widetilde{x}(f)\boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f df \right\|_2^2 \\
&= \left\| \int_{\mathbb{W}} \widetilde{x}(f)(\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f) df \right\|_2^2 \\
&\leq \int_{\mathbb{W}} |\widetilde{x}(f)|^2 df \cdot \int_{\mathbb{W}} \|\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f\|_2^2 df \\
&\leq \int_{\mathbb{W}} |\widetilde{x}(f)|^2 df \cdot C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W},\epsilon)N},
\end{aligned}$$

where the third line follows from the Cauchy-Schwarz inequality and the last line follows from (5.6) and the fact that $\int_{\mathbb{W}} \|\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f\|_2^2 df \leq |\mathbb{W}| \sup_{f\in\mathbb{W}} \|\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f\|_2^2 \leq \sup_{f\in\mathbb{W}} \|\boldsymbol{e}_f - \boldsymbol{P}_{\boldsymbol{\Psi}}\boldsymbol{e}_f\|_2^2$.

APPENDIX D

PROOFS FOR CHAPTER 6

This appendix contains the proofs for Chapter 6. Let $\operatorname{ess}\mathcal{R}(\cdot)$ be the essential range of a function. For any $\Omega \subset \mathbb{R}$, let $\operatorname{int}(\Omega)$ be the interior of the set Ω.

D.1 Proof of Theorem 6.2

We first provide a strong condition under which the equal distribution of two sequences is equivalent to individual asymptotic equivalence. Its proof is deferred to Appendix D.6.

Theorem D.1. *Assume that the sequences* $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ *and* $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ *are absolutely bounded , i.e., there exist* a', b' *such that* $b' \geq u_{N,0} \geq u_{N,1} \geq \cdots \geq u_{N,N-1} \geq a'$ *and* $b' \geq v_{N,0} \geq v_{N,1} \geq \cdots \geq v_{N,N-1} \geq a'$ *for all* $N \in \mathbb{N}$. *Furthermore, suppose there exists a non-constant continuous function* $g(x) : [c, d] \to \mathbb{R}$ *such that*

$$\lim_{N\to\infty} u_{N,0} = \lim_{N\to\infty} v_{N,0} = \max_{x\in[c,d]} g(x),$$
$$\lim_{N\to\infty} u_{N,N-1} = \lim_{N\to\infty} v_{N,N-1} = \min_{x\in[c,d]} g(x),$$

and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{l=0}^{N-1} \vartheta(u_{N,l}) = \frac{1}{d-c}\int_c^d \vartheta(g(x))dx < \infty$$

for every function ϑ *that is continuous on* $[a, b]$, *where [a,b] is the smallest interval that covers* $[a', b']$ *and the range of* $g(x)$. *Then the following are equivalent:*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{l=0}^{N-1} (\vartheta(u_{N,l}) - \vartheta(v_{N,l})) = 0; \tag{D.1}$$

$$\lim_{N\to\infty} \max_{l\in[N]} |u_{N,l} - v_{N,l}| = 0. \tag{D.2}$$

If $\widetilde{h}(f) \equiv C$ is a constant function, then $\boldsymbol{H}_N$ and $\boldsymbol{C}_N$ are diagonal matrices with all diagonals being C for all $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$ and $N \in \mathbb{N}$. Thus $\lambda_l\left(\boldsymbol{H}_N\right) = \lambda_l\left(\boldsymbol{C}_N\right) = C$ for all $l \in [N]$ and $\boldsymbol{C}_N \in \left\{\widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N\right\}$. The following result establishes the range of the eigenvalues of $\overline{\boldsymbol{C}}_N$ and $\boldsymbol{H}_N$ for the case when $\widetilde{h}(f)$ is not a constant function.

Lemma D.1. *Suppose that $\widetilde{h} \in L^\infty([0,1])$ and $\widetilde{h}$ is not a constant function. Also let $\lambda_l(\overline{\boldsymbol{C}}_N)$ be permuted such that $\lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N) \geq \lambda_{\rho(1)}(\overline{\boldsymbol{C}}_N) \geq \cdots \geq \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$. Then*

$$\operatorname{ess\,inf} \widetilde{h} < \lambda_{N-1}(\boldsymbol{H}_N) < \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$$

and

$$\lambda_{\rho(0)}(\overline{\boldsymbol{C}}_N) \leq \lambda_0(\boldsymbol{H}_N) < \operatorname{ess\,sup} \widetilde{h}.$$

Proof of Lemma D.1. We first rewrite $\lambda_l\left(\overline{\boldsymbol{C}}_N\right)$ as

$$\begin{aligned}
\lambda_l\left(\overline{\boldsymbol{C}}_N\right) &= \sum_{n=0}^{N-1} \overline{c}[n] e^{\frac{-j2\pi ln}{N}} \\
&= \sum_{n=0}^{N-1} \frac{1}{N}\left((N-n)h[-n] + nh[N-n]\right) e^{\frac{-j2\pi ln}{N}} \\
&= \left\langle \boldsymbol{H}_N \frac{1}{\sqrt{N}} \boldsymbol{e}_{l/N}, \frac{1}{\sqrt{N}} \boldsymbol{e}_{l/N} \right\rangle.
\end{aligned}$$

By definition, $\lambda_0(\boldsymbol{H}_N) = \max_{\|\boldsymbol{v}\|_2=1} \langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v} \rangle$ and $\lambda_{N-1}(\boldsymbol{H}_N) = \min_{\|\boldsymbol{v}\|_2=1} \langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v} \rangle$, we obtain

$$\lambda_{N-1}(\boldsymbol{H}_N) \leq \lambda_l(\overline{\boldsymbol{C}}_N) \leq \lambda_0(\boldsymbol{H}_N), \ \forall\ l.$$

For arbitrary $\boldsymbol{v} \in \mathbb{C}^N, \|\boldsymbol{v}\|_2 = 1$, we extend $\boldsymbol{v}$ to an infinite sequence $v[n], n \in \mathbb{Z}$ by zero-padding. Then

$$\begin{aligned}\langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v} \rangle &= \sum_{m=0}^{N-1} \boldsymbol{v}^*[m] \sum_{n=0}^{N-1} h[m-n] \boldsymbol{v}[n] \\ &= \sum_{m=-\infty}^{\infty} v^*[m] \sum_{n=-\infty}^{\infty} h[m-n] v[n] \\ &= \int_0^1 |\widetilde{\boldsymbol{v}}(f)|^2 \, \widetilde{h}(f) df\end{aligned}$$

where $\widetilde{\boldsymbol{v}}(f) = \sum_{n=0}^{N-1} \boldsymbol{v}[n] e^{j2\pi f n}$. If $\widetilde{h}(f)$ is not a constant function of $[0,1]$, we conclude

$$\begin{aligned}\operatorname{ess\,inf} \widetilde{h} &= \int_0^1 |\widetilde{\boldsymbol{v}}(f)|^2 \, df \cdot \operatorname{ess\,inf} \widetilde{h} < \langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v} \rangle \\ &< \int_0^1 |\widetilde{\boldsymbol{v}}(f)|^2 \, df \cdot \operatorname{ess\,sup} \widetilde{h} = \operatorname{ess\,sup} \widetilde{h}.\end{aligned}$$

□

Theorem 6.2 holds trivially when $\widetilde{h}(f)$ is a constant function since for this case $\boldsymbol{H}_N$ and $\boldsymbol{C}_N$ have the same eigenvalues for all $\boldsymbol{C}_N \in \left\{ \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N \right\}$ and $N \in \mathbb{N}$. In what follows, we suppose $\widetilde{h}(f)$ is not a constant function. The assumption of absolute summability of the sequence $h[k]$ indicates that its DTFT $\widetilde{h}(f)$ is continuous on $[0,1]$, and moreover, its partial Fourier sum $S_N(f)$ converges uniformly to $\widetilde{h}(f)$ on $[0,1]$ as $N \to \infty$ [77]. Thus, given $\epsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that

$$\left| \widetilde{h}(f) - S_{N-1}(f) \right| \leq \epsilon$$

for all $f \in [0,1]$ and $N \geq N_0$. The Cesàro sum $\sigma_N(f)$ also converges to $\widetilde{h}(f)$ uniformly on $[0,1]$ as $N \to \infty$.

Since the eigenvalues of $\widetilde{\boldsymbol{C}}_N$ and $\widehat{\boldsymbol{C}}_N$ are, respectively, the samples of $S_{N-1}(f)$ and $S_{\lfloor \frac{N-1}{2} \rfloor}(f)$, we conclude that $\widetilde{\boldsymbol{C}}_N$ and $\widehat{\boldsymbol{C}}_N$ are absolutely bounded. Lemma D.1 implies that $\overline{\boldsymbol{C}}_N$ and $\boldsymbol{H}_N$ are also absolutely bounded.

We next show $\lim_{N\to\infty} \max_l \lambda_l (\boldsymbol{C}_N) = \max_{f\in[0,1]} \widetilde{h}(f)$ for all $\boldsymbol{C}_N \in \left\{ \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N \right\}$. The extreme value theorem states that $\widetilde{h}(f)$ must attain a maximum and a minimum each at

least once since $\widetilde{h}(f) \in \mathbb{R}$ is continuous on $[0,1]$. Let

$$\widehat{f} := \arg\max_{f} \widetilde{h}(f)$$

denote any point at which $\widetilde{h}$ achieves its maximum value. Also let

$$\widehat{l}_N := \arg\min_{l \in [N]} \left| \widehat{f} - \frac{l}{N} \right|$$

denote any closest on-grid point to $\widehat{f}$. For arbitrary $\epsilon > 0$, by uniform convergence, there exists N_0 such that

$$\left| \lambda_{\widehat{l}_N} (\boldsymbol{C}_N) - \widetilde{h} \left(\frac{\widehat{l}_N}{N} \right) \right| \leq \epsilon$$

for all $N \geq N_0$. Noting that $\left| \frac{\widehat{l}_N}{N} - \widehat{f} \right| \leq \frac{1}{2N}$ and $\widetilde{h}$ is continuous on $[0,1]$, there exists $N_1 \in \mathbb{N}$ so that

$$\left| \widetilde{h} \left(\widehat{f} \right) - \widetilde{h} \left(\frac{\widehat{l}_N}{N} \right) \right| \leq \epsilon$$

when $N \geq N_1$. Thus we conclude

$$\left| \lambda_{\widehat{l}_N} (\boldsymbol{C}_N) - \widetilde{h} \left(\widehat{f} \right) \right| \leq 2\epsilon$$

for all $N \geq \max\{N_0, N_1\}$. Since ϵ is arbitrary,

$$\lim_{N \to \infty} \max_{l} \lambda_l (\boldsymbol{C}_N) = \max_{f \in [0,1]} \widetilde{h}(f)$$

for all $\boldsymbol{C}_N \in \left\{ \widetilde{\boldsymbol{C}}_N, \widehat{\boldsymbol{C}}_N, \overline{\boldsymbol{C}}_N \right\}$. Noting that $\lambda_l \left(\overline{\boldsymbol{C}}_N \right) \leq \lambda_0(\boldsymbol{H}_N) \leq \max_{f \in [0,1]} \widetilde{h}(f)$, we obtain

$$\lim_{N \to \infty} \max_{l} \lambda_l (\boldsymbol{H}_N) = \max_{f \in [0,1]} \widetilde{h}(f).$$

The asymptotic argument for the smallest eigenvalues can be obtained with a similar approach. It follows from Lemma 6.1 that $\boldsymbol{H}_N \sim \widehat{\boldsymbol{C}}_N \sim \widetilde{\boldsymbol{C}}_N \sim \overline{\boldsymbol{C}}_N$ and from Szegő's theorem (2.23) that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widetilde{h}(f))df.$$

Finally, the proof of Theorem 6.2 is completed by applying Theorem D.1 with $g = \widetilde{h}$.

D.2 Proof of Theorem 6.3

We first outline the main idea. Let $[\boldsymbol{H}_N]_{N-r}$ be the $(N-r)\times(N-r)$ matrix obtained by deleting the last r columns and the last r rows of $\boldsymbol{H}_N$. Similar notation holds for $[\widetilde{\boldsymbol{C}}_N]_{N-r}$. Note that $[\boldsymbol{H}]_{N-r}$ and $[\widetilde{\boldsymbol{C}}_N]_{N-r}$ have the same eigenvalues when $N > 2r$ since $[\boldsymbol{H}]_{N-r}$ is exactly the same as $[\widetilde{\boldsymbol{C}}_N]_{N-r}$. Also $\widehat{\boldsymbol{C}}_N$ is equivalent to $\widetilde{\boldsymbol{C}}_N$ when $N > 2r$. We first apply the Sturmian separation theorem for the Toeplitz and circulant matrices to obtain a bound on the distance between $\lambda_l(\boldsymbol{H}_N)$ and $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$. We then utilize the fact that $\widetilde{h}(f)$ is Lipschitz continuous to guarantee the closeness between $\lambda_l(\widetilde{\boldsymbol{C}}_N)$ and $\lambda_{l+r}(\widetilde{\boldsymbol{C}}_N)$. Finally, we show $\lambda_l(\widetilde{\boldsymbol{C}}_N)$ is close to $\lambda_l(\overline{\boldsymbol{C}}_N)$ since the Cesàro sum and partial Fourier sum converge to the same function in this case.

In order to prove Theorem 6.3, we establish the following useful results.

Lemma D.2. *Let $u_0, u_1, \ldots, u_{N-1} \in \mathbb{R}$ be an unordered sequence of N elements. We decreasingly arrange this sequence so that $u_{\rho(0)} \geq u_{\rho(1)} \geq \cdots \geq u_{\rho(N-1)}$. Then for any $r \in \{1, 2 \ldots, N-1\}$, we have*

$$\max_{1\leq r'\leq r} \max_{l\in[N-r'-1]} u_{\rho(l)} - u_{\rho(l+r')} \leq \max_{1\leq r'\leq r} \max_{l\in[N-r'-1]} |u_l - u_{l+r'}|.$$

Proof of Lemma D.2. The proof is straightforward for the case $r = 1$. If the sequence is constant, then

$$\max_{l\in[N-2]} u_{\rho(l)} - u_{\rho(l+1)} = \max_{l\in[N-2]} |u_l - u_{l+1}| = 0.$$

Suppose the sequence is not constant, i.e., there exist at least $l_1, l_2 \in [N]$ so that $u_{l_1} \neq u_{l_2}$. Let

$$l' = \arg\max_{l \in [N-2]} u_{\rho(l)} - u_{\rho(l+1)}$$

denote any point at which $u_{\rho(l)} - u_{\rho(l+1)}$ achieves its maximum. Search the sequence $\{u_l\}_{l\in[N]}$ to find $u_{l''}$ that is smaller than $u_{\rho(l')}$ and its index l'' is closest to $\rho(l')$. Thus

$$\max\left\{\left|u_{l''} - u_{l''+1}\right|, \left|u_{l''} - u_{l''-1}\right|\right\} \geq \max_{l\in[N-2]} u_{\rho(l)} - u_{\rho(l+1)}.$$

Suppose $r \geq 2$. Similarly, the proof for a constant sequence is obvious. Suppose the sequence is not constant. Let

$$\{l', r'\} = \arg\max_{1\leq r''\leq r} \max_{l\in[N-r''-1]} u_{\rho(l)} - u_{\rho(l+r'')}.$$

If there are several pairs $\{l', r'\}$ have the same values, we choose the one that r' has the smallest value. If $r' = 1$, the proof is similar to the case $r = 1$. We suppose $r' \geq 2$. Thus there exist at least r' elements that are smaller than $u_{\rho(l')}$ and only $r' - 1$ elements that are greater than $u_{\rho(l'+r')}$ and smaller than $u_{\rho(l')}$. Search the sequence $\{u_l\}_{l\in[N]}$ to find $u_{l''}$ that is smaller than $u_{\rho(l')}$ and its index l'' is the r'-th closest to $\rho(l')$. Without loss of generality, suppose $l'' < \rho(l')$.

If $u_{l''} \leq u_{\rho(l'+r')}$, we have

$$\max_{1\leq r''\leq r'} \left|u_{l''} - u_{l''+r''}\right| \geq u_{\rho(l')} - u_{\rho(l'+r')}$$

since there is at least one element in $\{u_l, l'' + 1 \leq l \leq l'' + r'\}$ that is greater than or equal to $u_{\rho(l')}$.

If $u_{l''} > u_{\rho(l'+r')}$, there exists $l'' \leq l''' \leq 2\rho(l') - l''$ such that $u_{l'''}$ is smaller than or equal to $u_{\rho(l'+r')}$ (otherwise, there are r' elements that are greater than $u_{\rho(l'+r')}$ and smaller than $u_{\rho(l')}$). Also near $u_{l'''}$, there must exist at least one element that is not smaller than $u_{\rho(l')}$.

Then

$$\max_{1\leq r''\leq r'} \max\left\{\left|u_{l'''}-u_{l'''+r''}\right|,\left|u_{l'''}-u_{l'''-r''}\right|\right\}\geq u_{\rho(l')}-u_{\rho(l'+r')}.$$

This completes the proof. □

In words, the largest error between the contiguous elements of a sequence is not magnified when the sequence is rearranged in decreasing (or increasing) order.

The following result establishes that the largest error between two sequences is not magnified when both of the sequences are rearranged in decreasing (or increasing) order.

Lemma D.3. *Let $u_0,\ldots,u_{N-1}\in\mathbb{R}$ and $v_0,\ldots,v_{N-1}\in\mathbb{R}$ be two unordered sequences of N elements. We decreasingly arrange these sequences so that $u_{\rho(0)}\geq u_{\rho(1)}\geq\cdots u_{\rho(N-1)}$ and $v_{\rho(0)}\geq v_{\rho(1)}\geq\cdots v_{\rho(N-1)}$. Then*

$$\max_{l\in[N-1]}\left|u_{\rho(l)}-v_{\rho(l)}\right|\leq\max_{l\in[N-1]}\left|u_l-v_l\right|.$$

Proof of Lemma D.3. Let

$$r'=\arg\max_{r\in[N-1]}\left|u_{\rho(r)}-v_{\rho(r)}\right|$$

denote any point at which $\left|u_{\rho(r)}-v_{\rho(r)}\right|$ achieves its maximum and let l' be the index of $u_{\rho(r')}$. Without loss of generality, we suppose $u_{\rho(r')}\geq v_{\rho(r')}$. If $v_{l'}\leq v_{\rho(r')}$, we have $u_{l'}-v_{l'}\geq u_{\rho(r')}-v_{\rho(r')}$. Otherwise suppose $v_{l'}>v_{\rho(r')}$, which implies $r'\geq 1$. Since there are only r' elements in $\{u_l\}_{l\in[N]}$ that are greater than $u_{\rho(r')}$ and r' elements in $\{v_l\}_{l\in[N]}$ that are greater than $v_{\rho(r')}$, there must exist l'' such that $u_{l''}\geq u_{\rho(r')}$ and $v_{l''}\leq v_{\rho(r')}$. Hence

$$u_{l''}-v_{l''}\geq u_{\rho(r')}-v_{\rho(r')}.$$

□

Lemma D.4. *[67, Sturmian separation theorem] Let $\boldsymbol{A}_N$ be an $N\times N$ Hermitian matrix and let $[\boldsymbol{A}_N]_{N-1}$ be the $(N-1)\times(N-1)$ matrix obtained by deleting the last column and the last*

row of $\boldsymbol{A}_N$. *Also let* $\lambda_0(\boldsymbol{A}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{A}_N)$ *and* $\lambda_0([\boldsymbol{A}_N]_{N-1}) \geq \cdots \geq \lambda_{N-2}([\boldsymbol{A}_N]_{N-1})$ *respectively denote the descending eigenvalues of* $\boldsymbol{A}_N$ *and* $[\boldsymbol{A}_N]_{N-1}$. *Then*

$$\lambda_l(\boldsymbol{A}_N) \geq \lambda_l([\boldsymbol{A}_N]_{N-1}) \geq \lambda_{l+1}(\boldsymbol{A}_N)$$

for all $0 \leq l \leq N-2$.

The above Sturmian separation theorem forms the foundation of the following analysis. We note that Zizler et.al. [150] utilized the Sturmian separation theorem to prove a refinement of Szegő's asymptotic formula in terms of the number of eigenvalues inside a given interval.

Now we are well equipped to prove Theorem 6.3. In what follows, we consider $N > 2r$. Note that in this case $\widetilde{\boldsymbol{C}}_N$ is equivalent to $\widehat{\boldsymbol{C}}_N$ and the eigenvalues of $\widetilde{\boldsymbol{C}}_N$ are the DFT samples of $S_{N-1}(f) = \widetilde{h}(f) = \sum_{k=-r}^{r} h[k]e^{j2\pi fk}$. Recall that $[\boldsymbol{H}_N]_{N-r}$ is the $(N-r)\times(N-r)$ matrix obtained by deleting the last r columns and the last r rows of $\boldsymbol{H}_N$. Similar notation holds for $[\widetilde{\boldsymbol{C}}_N]_{N-r}$.

Note that $[\boldsymbol{H}]_{N-r}$ is exactly the same as $[\widetilde{\boldsymbol{C}}_N]_{N-r}$ as they have the same elements when $N > 2r$. Thus $[\boldsymbol{H}]_{N-r}$ and $[\widetilde{\boldsymbol{C}}_N]_{N-r}$ have the same eigenvalues. Let $\lambda_l([\widetilde{\boldsymbol{C}}_N]_{N-r})$ be permuted such that

$$\lambda_{\rho(0)}([\widetilde{\boldsymbol{C}}_N]_{N-r}) \geq \cdots \geq \lambda_{\rho(N-r-1)}([\widetilde{\boldsymbol{C}}_N]_{N-r}).$$

We first consider the simple case when $r = 1$. It follows from the Sturmian separation theorem that

$$\begin{aligned}
&\lambda_l(\boldsymbol{H}_N) \geq \lambda_l([\boldsymbol{H}_N]_{N-1}) \geq \lambda_{l+1}(\boldsymbol{H}_N),\\
&\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N) \geq \lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-1}) \geq \lambda_{\rho(l+1)}(\widetilde{\boldsymbol{C}}_N)
\end{aligned}$$

for all $0 \leq l \leq N-2$. This implies the following relationship between $\lambda_l(\boldsymbol{H}_N)$ and $\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)$

$$\begin{aligned}
\lambda_l(\boldsymbol{H}_N) \leq &\lambda_{l-1}([\boldsymbol{H}_N]_{N-1}) = \lambda_{\rho(l-1)}([\widetilde{\boldsymbol{C}}_N]_{N-1}) \leq \lambda_{\rho(l-1)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l = 1, 2, \ldots, N-1,\\
\lambda_l(\boldsymbol{H}_N) \geq &\lambda_l([\boldsymbol{H}_N]_{N-1}) = \lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-1}) \geq \lambda_{\rho(l+1)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l = 0, 1, \ldots, N-2
\end{aligned} \tag{D.3}$$

which give

$$\begin{aligned}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| &= \max\left\{\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right), \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_l\left(\boldsymbol{H}_N\right)\right\} \\ &\leq \max\left\{\lambda_{\rho(l-1)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right), \lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l+1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\}\end{aligned}$$

for all $1 \leq l \leq N-2$. Applying Lemma D.2 with $r=1$, we obtain

$$\begin{aligned}\max_{1\leq l\leq N-2}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| &\leq \max_{0\leq l\leq N-2}\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l+1)}\left(\widetilde{\boldsymbol{C}}_N\right) \\ &\leq \max_{0\leq l\leq N-2}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{l+1}\left(\widetilde{\boldsymbol{C}}_N\right)\right|.\end{aligned}$$

Note that $\widetilde{h}(f)$ is Lipschitz continuous since it is continuously differentiable. There exists a Lipschitz constant K such that, for all f_1 and f_2 in $[0,1]$,

$$\left|\widetilde{h}(f_1)-\widetilde{h}(f_2)\right| \leq K\left|f_1-f_2\right|.$$

From the fact that the eigenvalues of $\widetilde{\boldsymbol{C}}_N$ are the DFT samples of $\widetilde{h}(f)$, i.e., $\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)=\widetilde{h}(\frac{l}{N})$, it follows that

$$\begin{aligned}\max_{1\leq l\leq N-2}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| &\leq \max_{0\leq l\leq N-2}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{l+1}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \\ &\leq \max_{0\leq l\leq N-2}\left|\widetilde{h}(\frac{l}{N})-\widetilde{h}(\frac{l+1}{N})\right| \leq K\frac{1}{N}.\end{aligned} \tag{D.4}$$

Utilizing the fact that $\lambda_0(\boldsymbol{H}_N) \leq \max_{f\in[0,1]}\widetilde{h}(f)$ and $\lambda_{N-1}(\boldsymbol{H}_N) \geq \min_{f\in[0,1]}\widetilde{h}(f)$ (see Lemma D.1) and applying (D.3) with $l=0$ which gives

$$\lambda_0\left(\boldsymbol{H}_N\right) \geq \lambda_{\rho(1)}\left(\widetilde{\boldsymbol{C}}_N\right),$$

we have

$$\begin{aligned}\left|\lambda_0\left(\boldsymbol{H}_N\right)-\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| &= \max\left\{\lambda_0\left(\boldsymbol{H}_N\right)-\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right), \lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_0\left(\boldsymbol{H}_N\right)\right\} \\ &\leq \max\left\{\max_{f\in[0,1]}\widetilde{h}(f)-\lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right), \lambda_{\rho(0)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\} \\ &\leq K\frac{1}{N},\end{aligned}$$

where the second inequality follows because $\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)$ are uniform samples of $\widetilde{h}(f)$ with grid size $\frac{1}{N}$. Similarly, we have

$$\begin{aligned}
&\left|\lambda_{N-1}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \\
&=\max\left\{\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{N-1}\left(\boldsymbol{H}_N\right), \lambda_{N-1}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\} \\
&\leq \max\left\{\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)-\min_{f\in[0,1]}\widetilde{h}(f), \lambda_{\rho(N-2)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(N-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\} \\
&\leq K\frac{1}{N}.
\end{aligned}$$

Along with (D.4), we conclude

$$\max_{0\leq l\leq N-1}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\leq K\frac{1}{N}.$$

Now we consider the case $r>1$. Repeatedly applying the Sturmian separation theorem r times yields

$$\begin{aligned}
&\lambda_l(\boldsymbol{H}_N)\geq\lambda_l([\boldsymbol{H}_N]_{N-r})\geq\lambda_{l+r}(\boldsymbol{H}_N), \\
&\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)\geq\lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-r})\geq\lambda_{\rho(l+r)}(\widetilde{\boldsymbol{C}}_N)
\end{aligned}$$

for all $0\leq l\leq N-r-1$. Noting that $[\boldsymbol{H}_N]_{N-r}$ is the same as $[\widetilde{\boldsymbol{C}}_N]_{N-r}$, we have

$$\begin{aligned}
\lambda_l(\boldsymbol{H}_N)\leq&\lambda_{l-r}([\boldsymbol{H}_N]_{N-r})=\lambda_{\rho(l-r)}([\widetilde{\boldsymbol{C}}_N]_{N-r})\leq\lambda_{\rho(l-r)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l=r,r+1,\ldots,N-1, \\
\lambda_l(\boldsymbol{H}_N)\geq&\lambda_l([\boldsymbol{H}_N]_{N-r})=\lambda_{\rho(l)}([\widetilde{\boldsymbol{C}}_N]_{N-r})\geq\lambda_{\rho(l+r)}(\widetilde{\boldsymbol{C}}_N),\ \forall\ l=0,1,\ldots,N-r-1
\end{aligned}$$

which give

$$
\begin{aligned}
&\max_{r\le l\le N-r-1}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \\
&=\max_{r\le l\le N-r-1}\max\left\{\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right),\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_l\left(\boldsymbol{H}_N\right)\right\} \\
&\le\max_{r\le l\le N-r-1}\max\left\{\lambda_{\rho(l-r)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right),\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l+r)}\left(\widetilde{\boldsymbol{C}}_N\right)\right\} \\
&\le\max_{0\le l\le N-r-1}\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(l+r)}\left(\widetilde{\boldsymbol{C}}_N\right) \\
&\le\max_{1\le r'\le r}\max_{0\le l\le N-r'-1}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{l+r'}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \\
&\le K\frac{r}{N},
\end{aligned}
$$

where the third inequality follows from Lemma D.2. Since

$$
\lambda_{r-1}(\boldsymbol{H}_N)\le\cdots\le\lambda_0(\boldsymbol{H}_N)\le\max_{f\in[0,1]}\widetilde{h}(f),
$$

we bound $\left|\lambda_{r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right|, r'\le r-1$ by considering the following two cases: if $\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\le\lambda_{r'}\left(\boldsymbol{H}_N\right)$, we have

$$
\lambda_{r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\le\max_{f\in[0,1]}\widetilde{h}(f)-\lambda_{\rho(r-1)}\left(\widetilde{\boldsymbol{C}}_N\right)\le K\frac{r}{N};
$$

if $\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)>\lambda_{r'}\left(\boldsymbol{H}_N\right)$, we have

$$
\begin{aligned}
\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{r'}\left(\boldsymbol{H}_N\right)\le&\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(r'+r)}\left(\widetilde{\boldsymbol{C}}_N\right) \\
&\le\max_{1\le r''\le r}\max_{0\le l\le N-r''-1}\left|\lambda_l\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{l+r''}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \\
&\le K\frac{r}{N}
\end{aligned}
$$

where the second line follows because $\lambda_{r'}\left(\boldsymbol{H}_N\right)\ge\lambda_{\rho(r'+r)}\left(\widetilde{\boldsymbol{C}}_N\right)$ and the third line follows from Lemma D.2. Thus we have

$$
\left|\lambda_{r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\le K\frac{r}{N}
$$

for all $0\le r'\le r-1$. Similarly,

$$
\begin{aligned}
&\left|\lambda_{N-r'}\left(\boldsymbol{H}_N\right)-\lambda_{\rho(N-r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right| \\
&\leq \max\left\{\lambda_{\rho(N-r')}\left(\widetilde{\boldsymbol{C}}_N\right)-\min_{f\in[0,1]}\widetilde{h}(f), \lambda_{\rho(N-r'-r)}\left(\widetilde{\boldsymbol{C}}_N\right)-\lambda_{\rho(N-r')}\left(\widetilde{\boldsymbol{C}}_N\right)\right\} \\
&\leq \max\left\{K\frac{r}{N}, K\frac{r}{N}\right\}=K\frac{r}{N},
\end{aligned}
$$

for all $1\leq r'\leq r$. Therefore,

$$
\max_{0\leq l\leq N-1}\left|\lambda_l\left(\boldsymbol{H}_N\right)-\lambda_{\rho(l)}\left(\widetilde{\boldsymbol{C}}_N\right)\right|\leq Kr\frac{1}{N}.
$$

for all $N>2r$.

Note that $S_{r+1}(f)=S_{r+2}(f)=\cdots=S_{N-1}(f)$ which gives

$$
\sigma_N(f)=\frac{\sum_{n=0}^{N-1}S_n(f)}{N}=\frac{\sum_{n=0}^{r}S_n(f)}{N}+\frac{N-r-1}{N}S_{r+1}(f).
$$

Thus

$$
\begin{aligned}
|\sigma_N(f)-S_{N-1}(f)| &= \left|\frac{\sum_{n=0}^{r}S_n(f)}{N}-\frac{r+1}{N}S_{r+1}(f)\right| \\
&= \left|\sum_{n=0}^{r}S_n(f)-(r+1)S_{r+1}(f)\right|\frac{1}{N} \\
&= O(\frac{1}{N})
\end{aligned}
$$

uniformly on $[0,1]$ as $N\to\infty$. Therefore,

$$
\max_{0\leq l\leq N-1}\left|\lambda_l(\overline{\boldsymbol{C}}_N)-\lambda_l(\widetilde{\boldsymbol{C}}_N)\right|=\max_{0\leq l\leq N-1}\left|\sigma_N(\frac{l}{N})-S_{N-1}(\frac{l}{N})\right|=O(\frac{1}{N})
$$

as $N\to\infty$. Finally,

$$
\begin{aligned}
&\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\
&=\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)-\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)+\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\
&\le\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)-\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)\right|+\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\
&\le\max_{0\le l\le N-1}\left|\lambda_l(\overline{\boldsymbol{C}}_N)-\lambda_l(\widetilde{\boldsymbol{C}}_N)\right|+\max_{0\le l\le N-1}\left|\lambda_{\rho(l)}(\widetilde{\boldsymbol{C}}_N)-\lambda_l(\boldsymbol{H}_N)\right| \\
&=O(\frac{1}{N})
\end{aligned}
$$

as $N\to\infty$, where the second inequality follows from Lemma D.3.

D.3 Proof of Theorem 6.4

We first provide another condition (which, informally speaking, is weaker than that in Theorem D.1) under which the equal distribution of two sequences implies individual asymptotic equivalence. Its proof is deferred to Appendix D.7.

Theorem D.2. *Assume that* $b\ge u_{N,0}\ge u_{N,1}\ge\cdots\ge u_{N,N-1}\ge a$ *and* $b\ge v_{N,0}\ge v_{N,1}\ge\cdots\ge v_{N,N-1}\ge a$. *Furthermore, suppose there is a Riemann integrable function* $g(x):[c,d]\to[a,b]$ *such that*

$$
u_{N,l},v_{N,l}\in\operatorname{int}\left(\operatorname{ess}\mathcal{R}(g)\right),\ \ \forall\ l\in[N],N\in\mathbb{N},
$$

and

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\vartheta(u_{N,l})=\frac{1}{d-c}\int_c^d\vartheta(g(x))dx<\infty
$$

for all ϑ *that are continuous on* $[a,b]$. *Then the following are equivalent:*

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta(u_{N,l})-\vartheta(v_{N,l})\right)=0; \tag{D.5}
$$

$$\lim_{N\to\infty} \max_{l} |u_{N,l} - v_{N,l}| = 0. \tag{D.6}$$

If $\widetilde{h}(f) \equiv C$ is a constant function, then $\lambda_l\left(\boldsymbol{H}_N\right) = \lambda_l\left(\overline{\boldsymbol{C}}_N\right) = C$ for all $l \in [N]$. Thus Theorem 6.4 holds trivially. On the other hand, suppose that $\widetilde{h} \in L^\infty([0,1])$ is not a constant function and the essential range of $\widetilde{h}$ is $\left[\operatorname{ess\,inf} \widetilde{h}, \operatorname{ess\,sup} \widetilde{h}\right]$. It follows from Lemma D.1 that $\lambda_l\left(\boldsymbol{H}_N\right), \lambda_l\left(\overline{\boldsymbol{C}}_N\right) \in \operatorname{int}\left(\mathcal{R}\left(\widetilde{h}\right)\right)$ for all $l \in [N]$ and $N \in \mathbb{N}$. Using Lemma 6.1 and Szegő's theorem (see (2.23)), the fact that $h[k]$ is square summable together with the fact that $\boldsymbol{H}_N, \overline{\boldsymbol{C}}_N$ are absolutely bounded imply

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widetilde{h}(f)) df,$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \left(\vartheta(\lambda_l(\boldsymbol{H}_N)) - \vartheta(\lambda_l(\overline{\boldsymbol{C}}_N))\right) = 0$$

for all ϑ that are continuous on $\left[\operatorname{ess\,inf} \widetilde{h}, \operatorname{ess\,sup} \widetilde{h}\right]$. Finally, (6.3) follows from Theorem D.2 with $g = \widetilde{h}$, $u_{N,l} = \lambda_l(\boldsymbol{H}_N)$ and $v_{N,l} = \lambda_{\rho(l)}(\overline{\boldsymbol{C}}_N)$. This completes the proof of Theorem 6.4.

Remark D.1. Theorem D.1 requires that g is continuous and that the extreme values of the sequences asymptotically converge to the extreme values of g (but meanwhile the extreme values of the sequences can be outside of the range of g). Theorem D.2 requires the sequences to be strictly inside the range of g.

D.4 Proof of Theorem 6.5

Lemma D.5. *Let* $D_N(f) := \frac{\sin(\pi N f)}{\sin(\pi f)}$ *denote the Dirichlet kernel. Fix* $0 < W < \frac{1}{2}$. *We have*

$$\int_0^1 |D_N(f)|^2 \, df = N, \ \ \forall \ \ N \in \mathbb{N},$$
$$\int_W^{1-W} |D_N(f)|^2 \, df = O(1), \ \ \textit{when} \ \ N \to \infty.$$

Proof of Lemma D.5. Noting that $D_N(f) = \frac{\sin(\pi N f)}{\sin(\pi f)} = \frac{e^{j2\pi fN}}{e^{j\pi f}} \sum_{n=0}^{N-1} e^{j2\pi fn}$, we have

$$|D_N(f)|^2 = \left| \sum_{n=0}^{N-1} e^{j2\pi fn} \right|^2 = \left(\sum_{n=0}^{N-1} e^{j2\pi fn} \right) \left(\sum_{m=0}^{N-1} e^{-j2\pi fm} \right) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} e^{j2\pi f(n-m)}.$$

It follows that

$$\int_0^1 |D_N(f)|^2 \, df = \int_0^1 \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} e^{j2\pi f(n-m)} df = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \int_0^1 e^{j2\pi f(n-m)} df = N.$$

Fix $0 < W < \frac{1}{2}$. For any $f \in [W, 1-W]$, $|D_N(f)|$ is bounded above by $\frac{1}{\sin(\pi W)}$. Therefore,

$$\int_W^{1-W} |D_N(f)|^2 \, df \leq \int_W^{1-W} \frac{1}{\sin^2(\pi W)} df \leq \frac{1}{\sin^2(\pi W)}.$$

□

Since $\widetilde{h}$ is bounded and Riemann integrable over $[0,1]$, it follows from the Riemann-Lebesgue theorem that $\widetilde{h}$ is continuous almost everywhere in $[0,1]$. Thus we can select $f_0 \in [0,1]$ and a positive number W such that

$$\left| \widetilde{h}(f) - \operatorname{ess\,sup} \widetilde{h} \right| \leq \frac{\epsilon}{4}$$

holds almost everywhere for $|f - f_0| \leq W$. For any $\boldsymbol{v} \in \mathbb{C}^N$, we have

$$\begin{aligned}
\langle \boldsymbol{H}_N \boldsymbol{v}, \boldsymbol{v} \rangle &= \int_0^1 |\widetilde{\boldsymbol{v}}(f)|^2 \, \widetilde{h}(f) df \\
&= \int_{f_0 - W}^{f_0 + W} |\widetilde{\boldsymbol{v}}(f)|^2 \, \widetilde{h}(f) df + \int_{\substack{f \in [0,1] \\ f \notin [f_0 - W, f_0 + W]}} |\widetilde{\boldsymbol{v}}(f)|^2 \, \widetilde{h}(f) df \\
&\geq \left(\operatorname{ess\,sup} \widetilde{h} - \frac{\epsilon}{4} \right) \int_{f_0 - W}^{f_0 + W} |\widetilde{\boldsymbol{v}}(f)|^2 \, \widetilde{d}f \\
&\quad - \max \left(\left| \operatorname{ess\,inf} \widetilde{h} \right|, \left| \operatorname{ess\,sup} \widetilde{h} \right| \right) \cdot \int_{\substack{f \in [0,1] \\ f \notin [f_0 - W, f_0 + W]}} |\widetilde{\boldsymbol{v}}(f)|^2 \, \widetilde{d}f.
\end{aligned} \tag{D.7}$$

The DTFT of $\boldsymbol{e}_{l/N}$ is

$$\widetilde{\boldsymbol{e}}_{l/N}(f) = \frac{e^{-j\pi N\left(f-\frac{l}{\sqrt{N}}\right)}}{e^{-j\pi\left(f-\frac{l}{N}\right)}} D_N(f-\frac{l}{N}).$$

Fix $\widetilde{h}$, ϵ and W. If $N \geq \frac{1}{W}$, there always exists l' such that $\left|\frac{l'}{N} - f_0\right| \leq \frac{W}{2}$. It follows from Lemma D.5 that

$$\int_{\frac{l'}{N}-\frac{W}{2}}^{\frac{l'}{N}+\frac{W}{2}} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l'/N}(f)\right|^2 df = 1 - \frac{1}{N}\int_{\frac{W}{2}}^{1-\frac{W}{2}} |D_N(f)|^2 \, df = 1 - o(1)$$

as $N \to \infty$. Note that $[\frac{l'}{N} - \frac{W}{2}, \frac{l'}{N} + \frac{W}{2}] \subset [f_0 - W, f_0 + W]$. Thus there exists $N_1 \in \mathbb{N}$ such that for all $N \geq \max\left\{N_1, \frac{1}{W}\right\}$

$$\begin{aligned}
&\int_{f_0-W}^{f_0+W} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l'/N}(f)\right|^2 \widetilde{d}f \geq 1 - \frac{\epsilon}{4\left|\operatorname{ess\,sup}\widetilde{h}\right|}, \\
&\int_{\substack{f\in[0,1] \\ f\notin[f_0-W,f_0+W]}} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l'/N}(f)\right|^2 \widetilde{d}f \leq \frac{\epsilon}{2\cdot\max\left(\left|\operatorname{ess\,inf}\widetilde{h}\right|, \left|\operatorname{ess\,sup}\widetilde{h}\right|\right)}.
\end{aligned} \tag{D.8}$$

Combining (D.7) and (D.8) yields

$$\begin{aligned}
\lambda_{l'}\left(\overline{\boldsymbol{C}}_N\right) =& \langle \boldsymbol{H}_N \frac{1}{\sqrt{N}}\boldsymbol{e}_{l'/N}, \frac{1}{\sqrt{N}}\boldsymbol{e}_{l'/N}\rangle \\
\geq& \left(\operatorname{ess\,sup}\widetilde{h} - \frac{\epsilon}{4}\right)\left(1 - \frac{\epsilon}{4\left|\operatorname{ess\,sup}\widetilde{h}\right|}\right) - \frac{\epsilon}{2} \\
\geq& \operatorname{ess\,sup}\widetilde{h} - \frac{\epsilon}{4} - \frac{\epsilon}{4} + \frac{\epsilon^2}{16\left|\operatorname{ess\,sup}\widetilde{h}\right|} - \frac{\epsilon}{2} \\
\geq& \operatorname{ess\,sup}\widetilde{h} - \epsilon
\end{aligned}$$

for all $N \geq \max\left\{N_1, \frac{1}{W}\right\}$. Noting that $\lambda_{l'}\left(\overline{\boldsymbol{C}}_N\right) \leq \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) \leq \operatorname{ess\,sup}\widetilde{h}$, we have

$$\left|\lambda_{\rho(0)}\left(\boldsymbol{C}_N\right) - \operatorname{ess\,sup}\widetilde{h}\right| \leq \epsilon$$

for all $N \geq N_0$. Since ϵ is arbitrary, we conclude

$$\lim_{N\to\infty} \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,sup} \widetilde{h}.$$

With a similar argument, we have

$$\lim_{N\to\infty} \lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right) = \operatorname{ess\,inf} \widetilde{h}.$$

Noting that $\lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) \leq \lambda_0\left(\boldsymbol{H}_N\right) \leq \operatorname{ess\,sup} \widetilde{h}$ and $\operatorname{ess\,inf} \widetilde{h} \leq \lambda_{N-1}(\boldsymbol{H}_N) < \lambda_{\rho(N-1)}(\overline{\boldsymbol{C}}_N)$ (see Lemma D.1), we complete the proof by noting that

$$\begin{aligned}
\lim_{N\to\infty} \lambda_{\rho(0)}\left(\overline{\boldsymbol{C}}_N\right) &= \lim_{N\to\infty} \lambda_0\left(\boldsymbol{H}_N\right) = \operatorname{ess\,sup} \widetilde{h} \\
\lim_{N\to\infty} \lambda_{\rho(N-1)}\left(\overline{\boldsymbol{C}}_N\right) &= \lim_{N\to\infty} \lambda_{N-1}\left(\boldsymbol{H}_N\right) = \operatorname{ess\,inf} \widetilde{h}.
\end{aligned}$$

Remark D.2. Theorem 6.5 works only for the circulant matrix $\overline{\boldsymbol{C}}_N$ and not $\widetilde{\boldsymbol{C}}_N$ or $\widehat{\boldsymbol{C}}_N$. This is closely related to the fact that the partial Cesàro sum has better convergence than the partial Fourier sum [77].

Remark D.3. Theorem 6.5 only requires $\widetilde{h}$ to be bounded and Riemann integrable, while Theorem 6.4 requires the range of $\widetilde{h}$ to be connected.

D.5 Proof of Lemma 6.1

It follows from the definition of $\widehat{\boldsymbol{C}}_N$ that

$$
\begin{aligned}
&\left\|\boldsymbol{H}_N - \widehat{\boldsymbol{C}}_N\right\|_F^2 \\
&= \sum_{k=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} k\left(|h[k] - h[-N+k]|^2 + |h[-k] - h[N-k]|^2\right) + \sum_{k=\left\lfloor \frac{N-1}{2} \right\rfloor + 1}^{\lfloor N/2 \rfloor} k\left(|h[k]|^2 + |h[-k]|^2\right) \\
&\leq \sum_{k=1}^{\lfloor N/2 \rfloor} 2k\left(|h[k]|^2 + |h[-k]|^2 + |h[N-k]|^2 + |h[k-N]|^2\right) \\
&\leq \sum_{k=1}^{N-1} 2k\left(|h[k]|^2 + |h[-k]|^2\right).
\end{aligned}
$$

Fix $\epsilon > 0$. By assumption that the sequence $h[k]$ is square summable, there exists N_0 such that

$$
\sum_{k=N_0}^{\infty} |h[k]|^2 + |h[-k]|^2 \leq \epsilon.
$$

Thus we have

$$
\begin{aligned}
&\frac{1}{N}\left\|\boldsymbol{H}_N - \widehat{\boldsymbol{C}}_N\right\|_F^2 \leq \frac{1}{N}\sum_{k=1}^{N_0-1} 2k\left(|h[k]|^2 + |h[-k]|^2\right) + \frac{1}{N}\sum_{k=N_0}^{N} 2k\left(|h[k]|^2 + |h[-k]|^2\right) \\
&\leq \frac{1}{N}\sum_{k=1}^{N_0-1} 2k\left(|h[k]|^2 + |h[-k]|^2\right) + 2\sum_{k=N_0}^{N}\left(|h[k]|^2 + |h[-k]|^2\right) \leq \epsilon + 2\epsilon = 3\epsilon
\end{aligned}
$$

when $N \geq \max\{N_0, N_1\}$ with $N_1 \geq \sum_{k=1}^{N_0-1} 2k\left(|h[k]|^2 + |h[-k]|^2\right)/\epsilon$. Since ϵ is arbitrary, we obtain

$$
\lim_{N\to\infty} \frac{1}{N}\left\|\boldsymbol{H}_N - \widehat{\boldsymbol{C}}_N\right\|_F^2 = 0.
$$

Noting that $\boldsymbol{H}_N$ and $\widehat{\boldsymbol{C}}_N$ are absolutely bounded by assumption, we conclude $\boldsymbol{H}_N \sim \widehat{\boldsymbol{C}}_N$. The proofs of $\boldsymbol{H}_N \sim \widetilde{\boldsymbol{C}}_N$ and $\boldsymbol{H}_N \sim \overline{\boldsymbol{C}}_N$ follow from the same approach.

D.6 Proof of Theorem D.1

Set

$$F_g(\alpha) := \frac{1}{d-c}\mu\left\{x \in [c,d] : g(x) \leq \alpha\right\}, \tag{D.9}$$
$$F_{u_N}(\alpha) := \frac{1}{N}\#\left\{l \in [N], u_{N,l} \leq \alpha\right\},$$
$$F_{v_N}(\alpha) := \frac{1}{N}\#\left\{l \in [N], v_{N,l} \leq \alpha\right\}.$$

Here, $\mu(E)$ is the Lebsegue measure of a subset $E \in \mathbb{R}$.

Definition 6.1 states that the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are asymptotically equally distributed if

$$\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right) = 0$$

for all ϑ that are continuous on $[a,b]$. Here $[a,b]$ is the smallest interval that covers the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$.

Trench [126] strengthens this definition by showing the following result.

Lemma D.6. *[126, Asymptotcially (absolutely) equal distribution] Assume that $b \geq u_{N,0} \geq u_{N,1} \geq \cdots \geq u_{N,N-1} \geq a$ and $b \geq v_{N,0} \geq v_{N,1} \geq \cdots \geq v_{N,N-1} \geq a$. The following are equivalent:*

1. $\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right) = 0$ *for all ϑ that are continuous on $[a,b]$;*
2. $\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\left|\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right| = 0$ *for all ϑ that are continuous on $[a,b]$.*

Here the sequences $\{\{u_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ and $\{\{v_{N,l}\}_{l\in[N]}\}_{N=1}^{\infty}$ are said to be *absolutely asymptotically equally distributed* [126] if

$$\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\left|\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right| = 0$$

for all ϑ that are continuous on $[a,b]$.

Viewing $g : [c,d] \to \mathbb{R}$ as a random variable, in probabilistic language, F_g is the cumulative distribution function (CDF) associated to g. Also F_{u_N} and F_{v_N} can be viewed as the

CDF of the discrete random variables $\boldsymbol{u}_N : \{0, 1, \ldots, N-1\} \to \mathbb{R}$ defined by $\boldsymbol{u}_N(l) = u_{N,l}$ and $\boldsymbol{v}_N : \{0, 1, \ldots, N-1\} \to \mathbb{R}$ defined by $\boldsymbol{v}_N(l) = v_{N,l}$, respectively. It is well known that the CDF of a random variable is right continuous and non-decreasing. The following result, known as the Portmanteau Lemma, gives a number of equal descriptions of weak convergence in terms of the CDF and the means of the random variables.

Lemma D.7. *[131, Portmanteau Lemma] The following are equivalent:*

1. $\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(u_{N,l}) = \frac{1}{d-c} \int_c^d \vartheta(g(x))dx$, *for all bounded, continuous functions* ϑ;

2. $\lim_{N\to\infty} F_{u_N}(\alpha) = F_g(\alpha)$ *for every point* α *at which* F_g *is continuous.*

Despite the fact that $F_g(\alpha)$ is right continuous and non-decreasing everywhere, some stronger results about $F_g(\alpha)$ can be obtained by utilizing the fact that g is continuous on $[c, d]$.

Lemma D.8. *Let* $F_g(\alpha)$ *be defined as in* (D.9). *Then* $F_g(\alpha)$ *is strictly increasing on* $\mathcal{R}(g)$, *i.e., for every* $\alpha \in int(\mathcal{R}(g))$, *there exists* $\epsilon > 0$ *such that, for each pair* (α_1, α_2) *satisfying*

$$\min_{x\in[c,d]} g(x) \leq \alpha - \epsilon < \alpha_1 < \alpha < \alpha_2 < \alpha + \epsilon \leq \max_{x\in[c,d]} g(x),$$

we have

$$F_g(\alpha_1) < F_g(\alpha) < F_g(\alpha_2).$$

Proof of Lemma D.8. Since $g(x) : [c, d] \to \mathbb{R}$ is continuous, there exists ϵ such that $(\alpha - \epsilon, \alpha + \epsilon) \subset \mathcal{R}(g)$ for $\alpha \in \text{int}\,(\mathcal{R}(g))$. Let α_1 be an arbitrary value such that $\alpha - \epsilon < \alpha_1 < \alpha$ and let $\alpha_1' = \frac{\alpha+\alpha_1}{2} \in \mathcal{R}(g)$. Noting that g is continuous, we have

$$\mu\left\{x \in [c, d] : |g(x) - \alpha_1'| < \frac{\alpha - \alpha_1}{2}\right\} > 0.$$

Thus, we obtain

$$\begin{aligned} F_g(\alpha) - F_g(\alpha_1) &= \frac{1}{d-c}\mu\left\{x \in [c,d] : \alpha_1 < g(x) \leq \alpha\right\} \\ &\geq \frac{1}{d-c}\mu\left\{x \in [c,d] : \alpha_1 < g(x) < \alpha\right\} > 0. \end{aligned}$$

Similarly, we have $F_g(\alpha) < F_g(\alpha_2)$ for $\alpha < \alpha_2 < \alpha + \epsilon$. □

We are now ready to prove the main part. First we show that (D.2) implies (D.1). Fix ϑ being some continuous function on $[a,b]$ and $\epsilon > 0$. The Weierstrass approximation theorem states that there exists a polynomial p on $[a,b]$ such that

$$|\vartheta(t) - p(t)| \leq \frac{\epsilon}{3}$$

for all $t \in [a,b]$. Since p is a polynomial, there exists a constant C such that

$$|p(t_2) - p(t_1)| \leq C\,|t_2 - t_1|$$

for any $a \leq t_1 \leq t_2 \leq b$. Also (D.2) implies that there exists an $N_0 \in \mathbb{N}$ such that

$$|u_{N,l} - v_{N,l}| \leq \frac{\epsilon}{3C}, \ \ \forall\ l \in [N]$$

for all $N \geq N_0$. Therefore, we have

$$\begin{aligned} &|\vartheta(u_{N,l}) - \vartheta(v_{N,l})| \\ &\leq |\vartheta(u_{N,l}) - p(u_{N,l})| + |p(u_{N,l}) - p(v_{N,l})| + |\vartheta(v_{N,l}) - p(v_{N,l})| \\ &\leq \frac{\epsilon}{3} + C\frac{\epsilon}{3C} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

for all $l \in [N]$ and $N \geq N_0$. Thus

$$\left|\frac{1}{N}\sum_{l=0}^{N-1}\left(\vartheta\left(u_{N,l}\right) - \vartheta\left(v_{N,l}\right)\right)\right| \leq \frac{1}{N}\sum_{l=0}^{N-1}|\vartheta(u_{N,l}) - \vartheta(v_{N,l})| \leq \epsilon$$

for all $N \geq N_0$. Since ϵ is arbitrary, this implies (D.1).

Now let us show that (D.1) implies (D.2). We prove the statement (D.1) $\Rightarrow$ (D.2) by contradiction. Suppose (D.2) is not true, i.e., there exists an increasing sequence $\{M_k\}_{k=1}^{\infty}$ and $\epsilon_1 > 0$ such that

$$\max_{l \in [M_k]} |u_{M_k,l} - v_{M_k,l}| \geq 2\epsilon_1$$

for all $k \geq 1$. Let $l_k = \arg\max_{l \in [M_k]} |u_{M_k,l} - v_{M_k,l}|$ denote any point at which $|u_{M_k,l} - v_{M_k,l}|$ achieves its maximum, which implies $|u_{M_k,l_k} - v_{M_k,l_k}| \geq 2\epsilon_1$. Without loss of generality, we suppose $u_{M_k,l_k} \leq v_{M_k,l_k}$, i.e., $u_{M_k,l_k} \leq v_{M_k,l_k} - 2\epsilon_1$.

1. Suppose $u_{M_k,l_k} \geq \max_{x \in [c,d]} g(x)$, which indicates $v_{M_k,l_k} \geq 2\epsilon_1 + \max_{x \in [c,d]} g(x)$. This contradicts the assumption that $\lim_{N \to \infty} v_{N,0} = \max_{x \in [c,d]} g(x)$.

2. Suppose $u_{M_k,l_k} < \max_{x \in [c,d]} g(x)$. By assumption that

   $$\lim_{N \to \infty} u_{N,N-1} = \lim_{N \to \infty} v_{N,N-1} = \min_{x \in [c,d]} g(x),$$

   there exist $k_0 \in \mathbb{N}$ and $\alpha_k \in \text{int}\,(\mathcal{R}(g))$ such that

   $$0 \leq \alpha_k - u_{M_k,l_k} < \frac{\epsilon_1}{2}$$

   and F_g is continuous at α_k for all $k \geq k_0$. Noting that F_g is right continuous everywhere and strictly increasing at α_k (which is shown in Lemma D.8), there exist $\epsilon_2, \epsilon_3 > 0$ such that $\epsilon_2 \leq \frac{\epsilon_1}{2}$, F_g is continuous at $\alpha_k + \epsilon_2$, and

   $$F_g(\alpha_k + \epsilon_2) = F_g(\alpha_k) + 3\epsilon_3. \tag{D.10}$$

   Lemma D.7 indicates that

   $$\lim_{M_k \to \infty} F_{u_{M_k}}(\alpha) = F_g(\alpha)$$

   for every point α at which F_g is continuous. Thus there exist $k_1 \in \mathbb{N}, k_1 \geq k_0$ such that

$$\begin{aligned} \left|F_{u_{M_k}}(\alpha_k) - F_g(\alpha_k)\right| &< \epsilon_3, \\ \left|F_{u_{M_k}}(\alpha_k + \epsilon_2) - F_g(\alpha_k + \epsilon_2)\right| &< \epsilon_3 \end{aligned} \tag{D.11}$$

for all $k \geq k_1$. Thus, we have

$$\begin{aligned} &F_{u_{M_k}}(\alpha_k + \epsilon_2) - F_{u_{M_k}}(\alpha_k) \\ &= F_{u_{M_k}}(\alpha_k + \epsilon_2) - F_g(\alpha_k + \epsilon_2) + F_g(\alpha_k + \epsilon_2) - F_g(\alpha_k) + F_g(\alpha_k) - F_{u_{M_k}}(\alpha_k) \\ &\geq F_g(\alpha_k + \epsilon_2) - F_g(\alpha_k) - \left|F_g(\alpha_k) - F_{u_{M_k}}(\alpha_k)\right| - \left|F_{u_{M_k}}(\alpha_k + \epsilon_2) - F_g(\alpha_k + \epsilon_2)\right| \\ &\geq 3\epsilon_3 - \epsilon_3 - \epsilon_3 = \epsilon_3 \end{aligned}$$

for all $k \geq k_1$, where the last line follows from (D.10) and (D.11). Noting that the above equation is equivalent to

$$\frac{1}{M_k} \# \left\{ l \in [M_k], \alpha_k < u_{M_k,l} \leq \alpha_k + \epsilon_2 \right\} \geq \epsilon_3,$$

we have

$$\begin{aligned} &\frac{1}{M_k} \# \left\{ l \in [M_k], u_{M_k,l_k} < u_{M_k,l} \leq \alpha_k + \epsilon_2 \right\} \\ &\geq \frac{1}{M_k} \# \left\{ l \in [M_k], \alpha_k < u_{M_k,l} \leq \alpha_k + \epsilon_2 \right\} \geq \epsilon_3. \end{aligned}$$

Thus, we obtain

$$0 \leq u_{M_k,l_k - \lceil \epsilon_3 M_k \rceil} - u_{M_k,l_k} \leq \alpha_k + \epsilon_2 - u_{M_k,l_k} \leq \epsilon_1,$$

which implies

$$v_{M_k,l_k} - u_{M_k,l_k - \lceil \epsilon_3 M_k \rceil} \geq v_{M_k,l_k} - u_{M_k,l_k} + u_{M_k,l_k} - u_{M_k,l_k - \lceil \epsilon_3 M_k \rceil} \geq 2\epsilon_1 - \epsilon_1 \geq \epsilon_1.$$

Now taking $\vartheta(t) = t$, we obtain

$$\frac{1}{M_k}\sum_{l=0}^{M_k-1}|\vartheta(u_{M_k,l})-\vartheta(v_{M_k,l})| \geq \frac{1}{M_k}\sum_{l=l_k-\lceil\epsilon_3 M_k\rceil}^{l_k}|\vartheta(u_{M_k,l})-\vartheta(v_{M_k,l})|$$
$$\geq \frac{1}{M_k}\sum_{l=l_k-\lceil\epsilon_3 M_k\rceil}^{l_k}|\vartheta(u_{M_k,l})-\vartheta(v_{M_k,l_k})| \geq \epsilon_3\epsilon_1 > 0$$

for all $k \geq k_1$. This contradicts Lemma D.6.

D.7 Proof of Theorem D.2

Theorem D.3. *(Riemann-Lebesgue theorem [5, Theorem 7.48]) The function $g(x) \in L^\infty([a,b])$ is Riemann integrable over $[a,b]$ if and only if it is continuous almost everywhere in $[a,b]$.*

Despite the fact that $F_g(\alpha)$ is right continuous and non-decreasing everywhere, some stronger results about $F_g(\alpha)$ can be obtained at some point α since $g(x)$ is Riemann integrable.

Lemma D.9. *Suppose $g(x) : [c,d] \to [a,b]$ is Riemann integrable and let $F_g(\alpha)$ be defined as in* (D.9)*. Then $F_g(\alpha)$ is strictly increasing at α if $\alpha \in \operatorname{int}(\operatorname{ess}\mathcal{R}(g))$, i.e., there exists $\epsilon > 0$ such that, for every pair (α_1, α_2) such that $\alpha - \epsilon < \alpha_1 < \alpha < \alpha_2 < \alpha + \epsilon$, $F_g(\alpha_1) < F_g(\alpha) < F_g(\alpha_2)$.*

Proof of Lemma D.9. Since $g(x) : [c,d] \to [a,b]$ is Riemann integrable and $\alpha \in \operatorname{int}(\operatorname{ess}\mathcal{R}(g))$, there exists ϵ such that $(\alpha-\epsilon, \alpha+\epsilon) \subset \operatorname{ess}\mathcal{R}(g)$. Let α_1 be an arbitrary value such that $\alpha - \epsilon < \alpha_1 < \alpha$ and let $\alpha_1' = \frac{\alpha+\alpha_1}{2} \in \operatorname{ess}\mathcal{R}(g)$. It follows from the definition of essential range that

$$\mu\left\{x \in [c,d] : |g(x) - \alpha_1'| < \frac{\alpha-\alpha_1}{2}\right\} > 0.$$

Thus, we obtain

$$\begin{aligned} F_g(\alpha) - F_g(\alpha_1) =& \frac{1}{d-c}\mu\{x \in [c,d] : \alpha_1 < g(x) \leq \alpha\} \\ \geq& \frac{1}{d-c}\mu\{x \in [c,d] : \alpha_1 < g(x) < \alpha\} > 0. \end{aligned}$$

Similarly, we have $F_g(\alpha) < F_g(\alpha_2)$ for $\alpha < \alpha_2 < \alpha + \epsilon$. □

We are now ready to prove the main part using the same approach that was used to prove Theorem D.1.

i. First, we show that (D.6) implies (D.5). This part is the same as those in Appendix D.6.

ii. Now let us show that (D.5) implies (D.6). We prove the statement (D.5) ⇒ (D.6) by contradiction. Suppose (D.6) is not true, i.e., there exists an increasing sequence $\{M_k\}_{k=1}^{\infty}$ and $\epsilon_1 > 0$ such that

$$\max_{l \in [M_k]} |u_{M_k,l} - v_{M_k,l}| \geq 2\epsilon_1, \ \forall \ k \geq 1.$$

Let $l_k = \arg\max_{l \in [M_k]} |u_{M_k,l} - v_{M_k,l}|$ denote any point at which $|u_{M_k,l} - v_{M_k,l}|$ achieves its maximum. This implies $|u_{M_k,l_k} - v_{M_k,l_k}| \geq 2\epsilon_1$. Without loss of generality, we suppose F_g is continuous at u_{M_k,l_k} and $u_{M_k,l_k} \leq v_{M_k,l_k}$, i.e., $u_{M_k,l_k} \leq v_{M_k,l_k} - 2\epsilon_1$. Otherwise, one can always pick a $\widehat{u}_{M_k,l_k}$ that is close enough to u_{M_k,l_k} and such that F_g is continuous at $\widehat{u}_{M_k,l_k}$ since F_g is continuous almost everywhere.

By assumption, $u_{M_k,l_k} \in \operatorname{int}(\operatorname{ess}\mathcal{R}(g))$. Noting that F_g is right continuous everywhere and strictly increasing at u_{M_k,l_k} (which is shown in Lemma D.9), there exist $\epsilon_2 > 0$ and $\epsilon_3 > 0$ such that $\epsilon_2 < \epsilon_1$, F_g is continuous at $u_{M_k,l_k} + \epsilon_2$, and

$$F_g(u_{M_k,l_k} + \epsilon_2) = F_g(u_{M_k,l_k}) + 3\epsilon_3. \tag{D.12}$$

Lemma D.7 indicates that

$$\lim_{M_k \to \infty} F_{u_{M_k}}(\alpha) = F_g(\alpha)$$

for every point α at which F_g is continuous. Thus there exists $k_0 \in \mathbb{N}$ such that

$$\begin{aligned} \left|F_{u_{M_k}}(u_{M_k,l_k}) - F_g(u_{M_k,l_k})\right| &\leq \epsilon_3, \\ \left|F_{u_{M_k}}(u_{M_k,l_k} + \epsilon_2) - F_g(u_{M_k,l_k} + \epsilon_2)\right| &\leq \epsilon_3 \end{aligned} \tag{D.13}$$

for all $k \geq k_0$. Thus, we have

$$
\begin{aligned}
&F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2)-F_{u_{M_k}}(u_{M_k,l_k})\\
&=\Big(F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2)-F_g(u_{M_k,l_k}+\epsilon_2)\Big)+(F_g(u_{M_k,l_k}+\epsilon_2)-F_g(u_{M_k,l_k}))\\
&\quad+\Big(F_g(u_{M_k,l_k})-F_{u_{M_k}}(u_{M_k,l_k})\Big)\\
&\geq F_g(u_{M_k,l_k}+\epsilon_2)-F_g(u_{M_k,l_k})-\Big|F_{u_{M_k}}(u_{M_k,l_k}+\epsilon_2)-F_g(u_{M_k,l_k}+\epsilon_2)\Big|\\
&\quad-\Big|F_g(u_{M_k,l_k})-F_{u_{M_k}}(u_{M_k,l_k})\Big|\\
&\geq 3\epsilon_3-\epsilon_3-\epsilon_3=\epsilon_3
\end{aligned}
$$

for all $k \geq k_0$, where the last line follows from (D.12) and (D.13). Note that the above equation is equivalent to

$$
\frac{1}{M_k}\#\left\{l\in[M_k], u_{M_k,l_k}<u_{M_k,l}\leq u_{M_k,l_k}+\epsilon_2\right\}\geq\epsilon_3.
$$

Then

$$
0\leq u_{M_k,l_k-\lceil\epsilon_3 M_k\rceil}-u_{M_k,l_k}\leq u_{M_k,l_k}+\epsilon_2-u_{M_k,l_k}=\epsilon_2,
$$

which implies

$$
\begin{aligned}
v_{M_k,l_k}-u_{M_k,l_k-\lceil\epsilon_3 M_k\rceil}&\geq v_{M_k,l_k}-u_{M_k,l_k}+u_{M_k,l_k}-u_{M_k,l_k-\lceil\epsilon_3 M_k\rceil}\\
&\geq 2\epsilon_1-\epsilon_2\geq 2\epsilon_1-\epsilon_1\geq\epsilon_1.
\end{aligned}
$$

Now taking $\vartheta(t)=t$, we obtain

$$
\begin{aligned}
&\frac{1}{M_k}\sum_{l=0}^{M_k-1}|\vartheta(u_{M_k,l})-\vartheta(v_{M_k,l})|\geq\frac{1}{M_k}\sum_{l=l_k-\lceil\epsilon_3 M_k\rceil}^{l_k}|\vartheta(u_{M_k,l})-\vartheta(v_{M_k,l})|\\
&\geq\frac{1}{M_k}\sum_{l=l_k-\lceil\epsilon_3 M_k\rceil}^{l_k}|\vartheta(u_{M_k,l})-\vartheta(v_{M_k,l_k})|\geq\epsilon_3\epsilon_1>0
\end{aligned}
$$

for all $k\geq k_1$. This contradicts Lemma D.6.

D.8 Proof of Lemma 6.4

The following result indicates that the main lobe of the Dirichlet kernel contains most of its energy.

Lemma D.10. *Let* $D_N(f) = \frac{\sin(\pi N f)}{\sin(\pi f)}$ *be the Dirichlet kernel. Then*

$$\int_0^{\frac{1}{N}} |D_N(f)|^2 \, df \geq 0.45N.$$

Proof of Lemma D.10. Noting that $|D_N(f)| = \frac{|\sin(\pi N f)|}{|\sin(\pi f)|} \geq \frac{|\sin(\pi N f)|}{|\pi f|}$, we have

$$\begin{aligned}
&\int_0^{\frac{1}{N}} |D_N(f)|^2 \, df \geq \int_0^{\frac{1}{N}} \left|\frac{\sin(\pi N f)}{\pi f}\right|^2 df = \frac{N}{\pi}\int_0^{\pi} \left|\frac{\sin(f)}{f}\right|^2 df = \frac{N}{\pi}\int_0^{\pi} \frac{\frac{1-\cos(2f)}{2}}{f^2} df \\
&= \frac{N}{2\pi}\int_0^{\pi} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}(2f)^{2k}}{(2k)! f^2} df = \frac{N}{\pi}\sum_{k=1}^{\infty} \frac{(-1)^{k+1}(2\pi)^{2k-1}}{(2k)!(2k-1)} \\
&\geq \frac{N}{\pi}\sum_{k=1}^{8} \frac{(-1)^{(k+1)}(2\pi)^{2k-1}}{(2k)!\,(2k-1)} \geq 0.45N,
\end{aligned}$$

where the third line follows from the common Taylor series $\cos(2f) = \sum_{k=0}^{\infty}(-1)^k \frac{(2f)^{2k}}{(2k)!}$. □

Suppose N is a multiple of 4. Note that

$$\lambda_l(\overline{\boldsymbol{C}}_N) = \int_0^1 \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}\right|^2 \widetilde{h}(f) df = \int_0^{\frac{1}{4}} \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}(f)\right|^2 df + \int_{\frac{3}{4}}^1 \left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}(f)\right|^2 df.$$

If $l = N/4$, $\left|\frac{1}{\sqrt{N}}\widetilde{\boldsymbol{e}}_{l/N}(f)\right| = \frac{1}{\sqrt{N}}\left|D_N(f - \frac{1}{4})\right|$. Thus

$$\begin{aligned}
\lambda_{N/4}(\overline{\boldsymbol{C}}_N) =& \frac{1}{N}\int_0^{\frac{1}{4}} \left|D_N(f - \frac{1}{4})\right|^2 df + \frac{1}{N}\int_{\frac{3}{4}}^1 \left|D_N(f - \frac{1}{4})\right|^2 df \\
=& \frac{1}{N}\int_0^{1/2} |D_N(f)|^2 \, df = \frac{1}{N}\frac{1}{2}\int_0^1 |D_N(f)|^2 \, df = \frac{1}{2}.
\end{aligned}$$

Similarly, we have $\lambda_{3N/4}(\overline{\boldsymbol{C}}_N) = \frac{1}{2}$.

Now for any $l \in [N],\ \frac{l}{N} \in [0, \frac{1}{4}) \cup (\frac{3}{4}, 1]$, the main lobe of $D_N(f - \frac{l}{N})$ is inside the interval $[0, \frac{1}{4}] \cup [\frac{3}{4}, 1]$. Thus

$$\begin{aligned}\lambda_l(\overline{\boldsymbol{C}}_N) =& \frac{1}{N} \int_0^{\frac{1}{4}} \left| D_N(f - \frac{l}{N}) \right|^2 df + \frac{1}{N} \int_{\frac{3}{4}}^1 \left| D_N(f - \frac{l}{N}) \right|^2 df \\ \geq& \frac{2}{N} \int_0^{\frac{1}{N}} |D_N(f)|^2 \, df \geq 0.9.\end{aligned}$$

Similarly, for any $l \in [N],\ \frac{l}{N} \in (\frac{1}{4}, \frac{3}{4})$, we have

$$\begin{aligned}\lambda_l(\overline{\boldsymbol{C}}_N) =& \frac{1}{N} \int_0^{\frac{1}{4}} \left| D_N(f - \frac{l}{N}) \right|^2 df + \frac{1}{N} \int_{\frac{3}{4}}^1 \left| D_N(f - \frac{l}{N}) \right|^2 df \\ \leq& 1 - \frac{2}{N} \int_0^{\frac{1}{N}} |D_N(f)|^2 \, df \leq 0.1.\end{aligned}$$

The proof is completed by noting that $0 \leq \lambda_l(\overline{\boldsymbol{C}}_N) \leq 1$ for all $l \in [N]$.

APPENDIX E

PROOFS FOR CHAPTER 7

This appendix contains the proofs for Chapter 7.

E.1 Proof of Proposition 7.1

Note that

$$\begin{aligned}
d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) &= \inf_{\mathbb{M}_n} \sup_{x \in \mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))} \inf_{y \in \mathbb{M}_n} \|x - y\|_{L_2(\mathbb{A})} \\
&= \inf_{\mathbb{M}_n} \sup_{\|\alpha\| \leq 1} \|\mathcal{A}\alpha - \mathcal{P}_{\mathbb{M}_n}\| \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \sup_{\|\alpha\| \leq 1} \frac{|\langle \mathcal{A}\alpha, z \rangle|}{\|z\|} \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \sup_{\|\alpha\| \leq 1} \frac{|\langle \alpha, \mathcal{A}^* z \rangle|}{\|z\|} \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \frac{\|\mathcal{A}^* z\|}{\|z\|} \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \frac{\sqrt{\langle \mathcal{A}\mathcal{A}^* z, z \rangle}}{\|z\|} \\
&= \sqrt{\lambda_n}
\end{aligned}$$

where the last line follows from the the Weyl-Courant minimax theorem.

E.2 Proof of Theorem 7.1

We first note that $\chi_\xi(0) = 1$ for all $\xi \in \widehat{\mathbb{G}}$. Thus, we have

$$\sum_\ell \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau, \mathbb{B}}) = \int_{\mathbb{A}} K_{\mathbb{B}}(0) \, \mathrm{d}\, h = |A| \int_{\mathbb{B}} \chi_\xi(0) \, \mathrm{d}\, \xi = |\mathbb{A}_\tau||\mathbb{B}|. \tag{E.1}$$

We write the operator $(\mathcal{T}_{\mathbb{A}} \mathcal{B}_{\mathbb{B}} \mathcal{T}_{\mathbb{A}})^2$ as

$$(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}x)(g) = \int_{\mathbb{A}} K_{\mathbb{B}}(\widetilde{h}^{-1}\circ g)\left(\int_{\mathbb{A}} K_{\mathbb{B}}(h^{-1}\circ\widetilde{h})x(h)\,\mathrm{d}\,h\right)\mathrm{d}\,\widetilde{h}$$
$$= \int_{\mathbb{A}}\left(\int_{\mathbb{A}} K_{\mathbb{B}}(\widetilde{h}^{-1}\circ g)K_{\mathbb{B}}(h^{-1}\circ\widetilde{h})\,\mathrm{d}\,\widetilde{h}\right)x(h)\,\mathrm{d}\,h.$$

Thus,

$$\sum_{\ell}\lambda_{\ell}^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \int_{\mathbb{A}}\int_{\mathbb{A}} K_{\mathbb{B}}(\widetilde{h}^{-1}\circ h)K_{\mathbb{B}}(h^{-1}\circ\widetilde{h})\,\mathrm{d}\,\widetilde{h}\,\mathrm{d}\,h$$
$$= \int_{\mathbb{A}}\int_{\mathbb{A}}\left|K_{\mathbb{B}}(h^{-1}\circ\widetilde{h})\right|^2\mathrm{d}\,\widetilde{h}\,\mathrm{d}\,h$$

where we use the fact that $K_{\mathbb{B}}(h^{-1}\circ g) = \int_{\mathbb{B}}\chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi = (\int_{\mathbb{B}}\chi_\xi(g^{-1}\circ h)\,\mathrm{d}\,\xi)^*$ since $\chi_\xi(-g) = \chi_\xi^*(g)$. Applying the change of variable $\widetilde{h} = h\circ\overline{h}$, we obtain

$$\sum_{\ell}\lambda_{\ell}^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \int_{\mathbb{A}}\int_{A-h}\left|K_{\mathbb{B}}(\overline{h})\right|^2\mathrm{d}\,\overline{h}\,\mathrm{d}\,h = \int_{\mathbb{A}}\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h)\,\mathrm{d}\,h, \tag{E.2}$$

where $\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h) = \int_{A-h}\left|K_{\mathbb{B}}(\overline{h})\right|^2\mathrm{d}\,\overline{h} \geq 0$. The function $\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h)$ is dominated as

$$\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h) \leq \int_{\mathbb{G}}\left|K_{\mathbb{B}}(\overline{h})\right|^2\mathrm{d}\,\overline{h} = \int_{\mathbb{G}}\left|\int_{\mathbb{B}}\chi_\xi(\overline{h})\,\mathrm{d}\,\xi\right|^2\mathrm{d}\,\overline{h} = \int_{\mathbb{B}}|\chi_\xi(\overline{h})|^2\,\mathrm{d}\,\xi = |\mathbb{B}|$$

where we use Parseval's theorem. On the other hand, we have

$$\lim_{\tau\to\infty}\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h) = \int_{\mathbb{G}}\left|\int_{\mathbb{B}}\chi_\xi(\overline{h})\,\mathrm{d}\,\xi\right|^2\mathrm{d}\,\overline{h} = |\mathbb{B}|$$

for all $h\in\mathbb{G}$. Then we have

$$\lim_{\tau\to\infty}\sum_{\ell}\lambda_{\ell}^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \int_{\mathbb{A}}|\mathbb{B}|\,\mathrm{d}\,h = |\mathbb{A}_\tau||\mathbb{B}|.$$

Thus, we have

$$\sum_{\ell}\lambda_{\ell}^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{A}_\tau||\mathbb{B}| - o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{E.3}$$

Subtracting (E.3) from (E.1) gives

$$\sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \left(1 - \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\right) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{E.4}$$

Utilizing the fact that $0 \leq \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq 1$, we have

$$\frac{1}{\epsilon(1-\epsilon)} \mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; (\epsilon, 1-\epsilon)) \leq \sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \left(1 - \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\right) = o(|\mathbb{A}_\tau||\mathbb{B}|).$$

On the other hand, (E.4) also implies that

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})<1-\epsilon} \epsilon \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) < \sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \left(1 - \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\right) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{E.5}$$

Plugging this term into (E.1) gives

$$\begin{aligned} |\mathbb{A}_\tau||\mathbb{B}| = \sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) &= \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) + \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \\ &= \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) + o(|\mathbb{A}_\tau||\mathbb{B}|). \end{aligned}$$

Similarly, plugging (E.5) into (E.3) gives

$$|\mathbb{A}_\tau||\mathbb{B}| = \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) + o(|\mathbb{A}_\tau||\mathbb{B}|).$$

Combining the above two equations and the fact that $\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq 1$, we have

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) - \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{E.6}$$

On one hand, combining (E.6) with

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) - \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} 1 - \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})$$

gives

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; [1-\epsilon, 1]) - \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} 1 - \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \geq o(|\mathbb{A}_\tau||\mathbb{B}|),$$

which further implies

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) \geq |\mathbb{A}_\tau||\mathbb{B}| - o(|\mathbb{A}_\tau||\mathbb{B}|).$$

On the other hand, using (E.6) and

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) - \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \geq (1-\epsilon) \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\leq 1-\epsilon} 1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}),$$

we also have

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) - \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} 1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq o(|\mathbb{A}_\tau||\mathbb{B}|),$$

which further implies

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) \leq |\mathbb{A}_\tau||\mathbb{B}| + o(|\mathbb{A}_\tau||\mathbb{B}|).$$

Thus we obtain

$$\lim_{\tau\to\infty} \frac{\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1])}{|\mathbb{A}_\tau|} = |\mathbb{B}|.$$

E.3 Proof of Theorem 7.2

We first recall the eigendecompostion of $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}} = \sum_{\ell\geq 0} \lambda_\ell u_\ell u_\ell^*$, where λ_ℓ is short for $\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})$. Utilizing the fact that $u_\ell, \ell = 0,1,\ldots$ is a complete orthonormal basis for $L_2(\mathbb{A}_\tau)$, we rewrite the function in (7.19):

$$\begin{aligned}
\|\chi_\xi(g) - \mathcal{P}_{\mathbb{M}_n}\chi_\xi(g)\|_{L_2(\mathbb{A}_\tau)}^2 &= \sum_\ell \left|\langle (I-\mathcal{P}_{\mathbb{M}_n})\chi_\xi(g), u_\ell(g)\rangle_{L_2(\mathbb{A}_\tau)}\right|^2 \\
&= \sum_\ell \left\langle \left\langle (I-\mathcal{P}_{\mathbb{M}_n})\chi_\xi(g)\chi_\xi^*(h), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)} \\
&= \sum_\ell \left\langle \left\langle (I-\mathcal{P}_{\mathbb{M}_n})\chi_\xi(h^{-1}\circ g), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)}
\end{aligned}$$

where the second equality utilized the fact that $\mathcal{P}_{\mathbb{M}_n}$ is the orthogonal projector onto the subspace $\mathbb{M}_n$, and $\sum_\ell \left\langle \left\langle (I-\mathcal{P}_{\mathbb{M}_n})\chi_\xi(g)\chi_\xi^*(h), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(h)\right\rangle_{L_2(\mathbb{A}_\tau)}$ is equivalent to the trace of $(I-\mathcal{P}_{\mathbb{M}_n})\chi_\xi(g)\chi_\xi^*(h)$. Plugging this equation into (7.19) gives

$$
\begin{aligned}
&\int_{\mathbb{B}} \|\chi_\xi(g) - \mathcal{P}_{\mathbb{M}_n}\chi_\xi(g)\|^2_{L_2(\mathbb{A}_\tau)}\,\mathrm{d}\,\xi \\
&= \int_{\mathbb{B}} \sum_\ell \left\langle \left\langle (I - \mathcal{P}_{\mathbb{M}_n})\chi_\xi(\theta^{-1}\circ g), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)} \mathrm{d}\,\xi \\
&= \sum_\ell \int_{\mathbb{B}} \left\langle \left\langle (I - \mathcal{P}_{\mathbb{M}_n})\chi_\xi(\theta^{-1}\circ g), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)} \mathrm{d}\,\xi \\
&= \sum_\ell \left\langle (I - \mathcal{P}_{\mathbb{M}_n})\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}} u_\ell, u_\ell\right\rangle_{L_2(\mathbb{A}_\tau)} \\
&= \sum_\ell \lambda_\ell \left\langle (I - \mathcal{P}_{\mathbb{M}_n})u_\ell, u_\ell\right\rangle_{L_2(\mathbb{A}_\tau)}
\end{aligned}
$$

where the second line follows from monotone convergence theorem (since each term inside the summation is nonnegative). Thus, we conclude that the optimal n-dimensional subspace which minimizes the last term in the above equation is $\mathbb{U}_n$ (which is spanned by the first n eigenfunctions). With this choice of subspace, we have

$$
\begin{aligned}
\int_{\mathbb{B}} \|\chi_\xi(g) - \mathcal{P}_{\mathbb{U}_n}\chi_\xi(g)\|^2_{L_2(\mathbb{A}_\tau)}\,\mathrm{d}\,\xi &= \sum_{\ell\geq n} \lambda_\ell \\
&= |\mathbb{A}_\tau||\mathbb{B}| - \sum_{\ell=0}^{n-1} \lambda_\ell
\end{aligned}
$$

since by (7.8) we have $\sum_\ell \lambda_\ell = |\mathbb{A}_\tau||\mathbb{B}|$. The proof is completed by noting that $\|\chi_\xi(g)\|^2_{L_2(\mathbb{A}_\tau)} = |\mathbb{A}_\tau|$ for any $\xi \in \mathbb{B}$.

E.4 Proof of Theorem 7.3

First let ν be a random variable with uniform distribution on $[0, 2\pi)$. We define the random vector

$$
r(g) = r(g;\xi,\nu) = \chi_\xi(g)e^{j\nu},
$$

where the term $e^{j\nu}$ acts as a phase randomizer and ensures that r is zero-mean:

$$
\mathrm{E}\,[r(g)] = \frac{1}{|\mathbb{B}|2\pi}\int_{\mathbb{B}} \chi_\xi(g)e^{j\nu}\,\mathrm{d}\,\xi\,\mathrm{d}\,\nu = \frac{1}{|\mathbb{B}|2\pi}\int_{\mathbb{B}} \chi_\xi(g)\,\mathrm{d}\,\xi \int_0^{2\pi} e^{j\nu}\,\mathrm{d}\,\nu = 0
$$

for all $g \in \mathbb{A}_\tau$.

Now we compute the autocorrelation R of the random variable r as

$$\begin{aligned} R(g,h) &= \mathrm{E}\left[r(g)r^*(h)\right] \\ &= \mathrm{E}\left[\left(\chi_\xi(g)e^{j\nu}\right)\left(\chi_\xi^*(h)e^{-j\nu}\right)\right] \\ &= \mathrm{E}\left[\chi_\xi(\theta^{-1}\circ g)\right] \\ &= \frac{1}{|\mathbb{B}|}\int_{\mathbb{B}} \chi_\xi(\theta^{-1}\circ g)\,\mathrm{d}\,\xi \\ &= \frac{1}{|\mathbb{B}|}K_{\mathbb{B}}(\theta^{-1}\circ g) \end{aligned} \tag{E.7}$$

for all $h, g \in \mathbb{A}_\tau$. Here $K_{\mathbb{B}}$ is defined in (7.5). Note that $K_{\mathbb{B}}(\theta^{-1}\circ g)$ with $h, g \in \mathbb{A}_\tau$ is the kernel of the Toeplitz operator $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$. Now it follows from the Karhunen-Loève (KL) transfrom [121] that

$$\mathrm{E}\left[\|r - \mathcal{P}_{\mathbb{U}_n} r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \frac{1}{|\mathbb{B}|}\sum_{\ell \geq n} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{B}| - \sum_{\ell=0}^{n-1} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}).$$

We then compute the expectation for the energy of r as

$$\mathrm{E}\left[\|r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \frac{1}{|\mathbb{B}|}\frac{1}{2\pi}\int_{\mathbb{B}} |\chi_\xi(g)e^{j\nu}|^2\,\mathrm{d}\,\xi\,\mathrm{d}\,\nu = |\mathbb{A}_\tau|.$$

The proof is completed by noting that $\mathrm{E}\left[\|r - \mathcal{P}_{\mathbb{U}_n} r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \mathrm{E}\left[\|x - \mathcal{P}_{\mathbb{U}_n} x\|^2_{L_2(\mathbb{A}_\tau)}\right]$ and $\mathrm{E}\left[\|r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A}_\tau)}\right]$.

E.5 Proof of Corollary 7.1

First, the inverse Fourier transform of the power spectrum $P_x(\xi)$ gives the autocorrelation function for $x(g)$:

$$a_x(g) = \int_\xi \frac{1}{|\mathbb{B}|}\chi_\xi(g)\,\mathrm{d}\,\xi = \frac{1}{|\mathbb{B}|}K_{\mathbb{B}}(g).$$

It follows that the random vector x has mean zero and an autocorrelation function R given by (E.7). Thus, x has exactly the same autocorrelation structure as the random function r we considered in Appendix E.4. The proof is completed by computing

$$\mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A}_\tau)}\right] = \int_{\mathbb{A}_\tau} \frac{1}{|\mathbb{B}|}K_{\mathbb{B}}(e)\,\mathrm{d}\,g = |\mathbb{A}_\tau|.$$