

# Approximating Sampled Sinusoids and Multiband Signals Using Multiband Modulated DPSS Dictionaries

Zhihui Zhu and Michael B. Wakin\*

Department of Electrical Engineering and Computer Science  
Colorado School of Mines

## Abstract

Many signal processing problems—such as analysis, compression, denoising, and reconstruction—can be facilitated by expressing the signal as a linear combination of atoms from a well-chosen dictionary. In this paper, we study possible dictionaries for representing the discrete vector one obtains when collecting a finite set of uniform samples from a multiband analog signal. By analyzing the spectrum of combined time- and multiband-limiting operations in the discrete-time domain, 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 an 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.

**Keywords.** Multiband signals, Discrete Prolate Spheroidal Sequences, discrete Fourier transform, sampling, approximation, signal recovery

## 1 Introduction

### 1.1 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.

Throughout this paper, we consider the signal space  $\mathbb{C}^N$ , and we represent a dictionary as an  $N \times L$  matrix  $\Psi$ , which has columns (or atoms)  $\psi_0, \psi_1, \dots, \psi_{L-1}$ . Using this dictionary, a signal  $\mathbf{x} \in \mathbb{C}^N$  can be represented exactly or approximately as a linear combination of the  $\psi_i$ :

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

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:

$$\mathbf{x} \approx \sum_{i \in \Omega} \alpha[i] \psi_i, \tag{1}$$

---

\*Email: zzhu,mwakin@mines.edu. This work was supported by NSF grant CCF-1409261.

where  $\Omega \subset \{0, 1, \dots, L-1\}$  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

$$\mathbf{x} \approx \sum_{i \in \Omega(\mathbf{x})} \alpha[i] \psi_i,$$

where  $\Omega(\mathbf{x}) \subset \{0, 1, \dots, L-1\}$  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 [14, 15, 29]. Sparse approximations have been widely used for signal denoising [16], signal recovery [4] and compressive sensing (CS) [5, 6, 8, 10, 17], 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  $\mathbf{x}$  is to identify which of the  $\binom{L}{J}$  subspaces (or, equivalently, index sets  $\Omega(\mathbf{x})$ ) to use. 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 [5, 9, 18] and greedy algorithms [3, 30, 31, 40] provide suitable approximations.

## 1.2 Dictionaries for finite-length vectors of sampled analog signals

In this paper, we 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

$$\mathbf{x} = [x(0) \ x(T_s) \ \dots \ x((N-1)T_s)]^T \quad (2)$$

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  $\Psi$  that provides highly accurate approximations of  $\mathbf{x}$  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  $\mathbf{x}$ . 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 describe this model more fully in Section 1.2.2. Before doing so, we begin in Section 1.2.1 with a simpler analog signal model for which an efficient dictionary  $\Psi$  is easier to describe.

### 1.2.1 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  $\mathbf{x}$ , as defined in (2), 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

$$\mathbf{x}[n] = \sum_{i=0}^{J-1} \beta_i e^{j2\pi f_i n}, \quad n = 0, 1, \dots, N-1, \quad (3)$$

where  $f_i = F_i T_s$ . This model arises in problems such as radar signal processing with point targets [27] and super-resolution [7].

In certain cases, an effective dictionary for representing  $\mathbf{x}$  is the  $N \times N$  discrete Fourier transform (DFT) matrix [2, 41, 27], where  $\psi_i[n] = e^{j2\pi i n/N}$  for  $i = 0, 1, \dots, N-1$  and  $n = 0, 1, \dots, N-1$ . Using this dictionary, we can write  $\mathbf{x} = \Psi \boldsymbol{\alpha}$ , where  $\boldsymbol{\alpha} \in \mathbb{C}^N$  contains the DFT coefficients of  $\mathbf{x}$ . When the frequencies  $f_i$  appearing in (3) 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  $\mathbf{x}$  in  $\mathbb{C}^N$ . This ‘‘on grid’’ multitone signal is sometimes assumed for simplicity in the CS literature [41]. 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 [41], or iteratively using a

refined dictionary [20]. An alternative is to consider the model (3) directly as in [19, 39]. However, such approaches cannot be generalized to scenarios in which the analog signals contain several bands, each with non-negligible bandwidth.

### 1.2.2 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  $\{0, 1, \dots, J-1\}$ .

Again we let  $\mathbf{x}$ , as defined in (2), 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  $\mathbf{x}$  can be expressed as an integral of sampled pure tones (i.e., discrete-time sinusoids)

$$\mathbf{x}[n] = \int_{\mathbb{W}} \tilde{x}(f) e^{j2\pi f n} df, \quad n = 0, 1, \dots, N-1, \quad (4)$$

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

$$\mathbb{W} := T_s \mathbb{F} = [f_0 - W_0, f_0 + W_0] \cup [f_1 - W_1, f_1 + W_1] \cup \dots \cup [f_{J-1} - W_{J-1}, f_{J-1} + W_{J-1}] \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right] \quad (5)$$

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  $\tilde{x}(f)$  appearing in (4) equals the scaled CTFT of  $x(t)$ ,

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

and corresponds to the discrete-time Fourier transform (DTFT) of the infinite sample sequence  $\{\dots, x(-2T_s), x(-T_s), x(0), x(T_s), x(2T_s), \dots\}$ . (However, we stress that our interest is on the finite-length sample vector  $\mathbf{x}$  and not on this infinite sample sequence.) Such multiband signal models arise in problems such as radar signal processing with non-point targets [1] and mitigation of narrowband interference [11, 12].

In this paper, we focus on building a dictionary in which finite-length sample vectors arising from multiband analog signals can be well-approximated using a small number of atoms. The DFT basis is inefficient for representing these signals because the DFT frequencies comprise only a regular, finite grid rather than a continuum of frequencies as appears in (5). Consequently, as previously discussed, any ‘‘off grid’’ frequency content in  $x(t)$  will spread across the DFT frequencies when the signal is sampled and time-limited.

In the simplified case of a baseband signal model (where  $J = 1$ ,  $F_0 = 0$ , and  $T_s \ll \frac{1}{B_{\text{nyq}}}$ ), an efficient alternative to the DFT basis is given by the dictionary of Discrete Prolate Spheroidal Sequences (DPSS’s) [37]. DPSS’s are a collection of bandlimited sequences that are most concentrated in time to a given index range and the DPSS vectors are the finite-support sequences (or vectors) whose DTFT is most concentrated in a given bandwidth [37]; we review properties of DPSS’s in Section 2.3. DPSS’s provide a highly efficient basis for representing sampled bandlimited signals (when  $\mathbb{W}$  reduces to a simple band  $[-W_0, W_0]$ ) and have proved to be useful in numerous signal processing applications. For instance, extrapolating a signal from a finite set of samples is an important problem with applications in remote sensing and other areas [33]. One can apply DPSS’s to find the minimum energy, infinite-length bandlimited sequence that extrapolates a given finite vector of samples [37]. Another problem involves estimating time-varying channels in wireless communication systems. In [42], Zemen

and Mecklenbräuke showed that expressing the time-varying subcarrier coefficients with a DPSS basis yields better estimates than those obtained with a DFT basis, which suffers from frequency leakage.

By modulating the baseband DPSS vectors to different frequency bands and then concatenating these dictionaries, one can construct a new dictionary that provides an efficient representation of sampled multiband signals. Sejdíć et al. [36] proposed one such dictionary to provide a sparse representation for fading channels and improve channel estimation accuracy. Zemen et al. [43, 44] utilized multiband DPSS sequences for band-limited prediction and estimation of time-variant channels. In CS, Davenport and Wakin [13] studied multiband modulated DPSS dictionaries for recovery of sampled multiband signals, and Sejdíć et al. [35] applied these dictionaries for the recovery of physiological signals from compressive measurements. Ahmad et al. [1] used such dictionaries for mitigating wall clutter in through-the-wall radar imaging, and modulated DPSS's can also be useful for detecting targets behind the wall [45].

In most of these works, the dictionary is assembled by partitioning the digital bandwidth  $[-\frac{1}{2}, \frac{1}{2}]$  uniformly into many bands and constructing a modulated DPSS basis for each band. The key fact that makes such a dictionary useful is that finite-length sample vectors arising from multiband analog signals will tend to have a block-sparse representation in this dictionary, where only those bands in the dictionary overlapping the frequencies  $\mathbb{W}$  are utilized. With this block-sparse structure, [13] provided theoretical guarantees into the use of this dictionary for sparsely representing sampled multiband signals and recovering sampled multiband signals from compressive measurements.

To date, however, relatively little work has focused on providing formal approximation guarantees for sampled multiband signals using multiband modulated DPSS dictionaries. To the best of our knowledge, an approximation guarantee in a mean-square error (MSE) sense was first presented formally in [13]. However, the question of how this dictionary compares to an optimal one has not been addressed. The objective of this paper is to answer this question and related ones.

### 1.3 Contributions

We study multiband modulated DPSS dictionaries in terms of the subspaces they span on the respective bands. More specifically, let

$$\mathbf{e}_f := \begin{bmatrix} e^{j2\pi f0} \\ e^{j2\pi f1} \\ \vdots \\ e^{j2\pi f(N-1)} \end{bmatrix} \in \mathbb{C}^N, \quad f \in [-\frac{1}{2}, \frac{1}{2}]$$

denote a length- $N$  vector of samples from a discrete-time complex exponential signal with digital frequency  $f$ . Then, it follows directly from (4) that a multiband sample vector  $\mathbf{x}$  can be expressed as

$$\mathbf{x} = \int_{\mathbb{W}} \tilde{x}(f) \mathbf{e}_f df, \quad (6)$$

where  $\mathbb{W}$  is as defined in (5). We can interpret this equation geometrically: the sampled complex exponentials  $\{\mathbf{e}_f\}_{f \in [-1/2, 1/2]}$  comprise a one-dimensional submanifold of  $\mathbb{C}^N$ . The vectors

$$\mathcal{M}_{\mathbb{W}} := \{\mathbf{e}_f\}_{f \in \mathbb{W}}$$

trace out a union of  $J$  finite-length curves belonging to this manifold. The sample vector  $\mathbf{x}$  can be expressed as an integral over the vectors in  $\mathcal{M}_{\mathbb{W}}$ , with weights determined by  $\tilde{x}(f)$ .

We are interested in several questions relating to the union of curves  $\mathcal{M}_{\mathbb{W}}$ :

- What is its effective dimensionality? That is, what dimensionality of a union of subspaces could nearly capture the energy of all signals in  $\mathcal{M}_{\mathbb{W}}$ , in the  $\ell_2$  metric?
- What is a suitable basis for the collective span of this union of subspaces?

Since we consider  $\ell_2$  approximation error, we will approach the approximation problem via the Karhunen-Loève (KL) transform (also known as principal component analysis (PCA) [26]) [38, 13]. We can imagine drawing a vector randomly from  $\mathcal{M}_{\mathbb{W}}$  with random phase, and we study the covariance structure of this random vector. Its covariance matrix is  $\mathbf{B}_{N, \mathbb{W}}$ , which has entries

$$\mathbf{B}_{N, \mathbb{W}}[m, n] := \int_{\mathbb{W}} e^{j2\pi f(m-n)} df = \sum_{i=0}^{J-1} e^{j2\pi f_i(m-n)} \frac{\sin(2\pi W_i(m-n))}{\pi(m-n)} \quad (7)$$

for all  $m, n \in [N]$ . The eigen-decomposition of  $\mathbf{B}_{N, \mathbb{W}}$  provides the optimal dictionary for linearly approximating this random vector. In particular, the  $k$  eigenvectors corresponding to the  $k$  largest eigenvalues of  $\mathbf{B}_{N, \mathbb{W}}$  span the  $k$ -dimensional subspace of  $\mathbb{C}^N$  that best captures these random vectors in terms of MSE; the resulting MSE equals the sum of the  $N - k$  smallest eigenvalues. When  $k$  can be chosen such that this residual sum is indeed small, this indicates that the effective dimensionality (informally, the “information level”) of the set  $\mathcal{M}_{\mathbb{W}}$  is roughly  $k$ .

The first contribution of this paper is to investigate the spectrum of the matrix  $\mathbf{B}_{N, \mathbb{W}}$ , which is equivalent<sup>1</sup> to a composed time- and multiband- limiting operator  $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$  defined in Section 2.2. In line with analogous results for time-frequency localization in the continuous-time domain [25, 28], we extend some of the techniques from [25, 28] for the discrete-time case and show that the number of dominant eigenvalues of  $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$  (and hence  $\mathbf{B}_{N, \mathbb{W}}$ ) is essentially the time-frequency area  $N|\mathbb{W}| = \sum_i 2NW_i$ , which also reveals the effective dimensionality of the union of curves  $\mathcal{M}_{\mathbb{W}}$ . Furthermore, similar to the concentration behavior of the DPSS eigenvalues for a single frequency band, we show that the eigenvalues of the operator  $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$  have a distinctive behavior: the first  $\approx N|\mathbb{W}|$  eigenvalues tend to cluster near 1, while the remaining eigenvalues tend to cluster near 0. All of these facts tell us that  $\approx N|\mathbb{W}|$  atoms are needed in order to accurately approximate, in an MSE sense, discrete-time sinusoids with frequencies in  $\mathbb{W}$ . As indicated in (6), such discrete-time sinusoids are themselves the building blocks of sampled multiband signals.

The second contribution of this paper is to show that the multiband modulated DPSS dictionary is approximately the optimal one for representing sampled multiband signals. Specifically, we show that there is a near nesting relationship between the subspaces spanned by the true eigenvectors of  $\mathbf{B}_{N, \mathbb{W}}$  and by the multiband modulated DPSS vectors on the bands of interest.<sup>2</sup> Directly computing both baseband DPSS vectors and the eigenvectors of  $\mathbf{B}_{N, \mathbb{W}}$  can be difficult, as the clustering of the eigenvalues makes the problem ill-conditioned. However, several references such as [21, 37] have pointed out that the baseband DPSS’s can also be computed by noting that the corresponding prolate matrix commutes with a well-conditioned symmetric tridiagonal matrix. Thus, the multiband modulated DPSS dictionary, which merely consists of various modulations of baseband DPSS’s, can be constructed more easily than the optimal one (which consists of the eigenvectors of  $\mathbf{B}_{N, \mathbb{W}}$ ).

The third contribution of this paper is to confirm that the multiband modulated DPSS dictionary provides a high degree of approximation for *all* sample vectors  $\mathbf{e}_f$  of discrete-time sinusoids with frequencies  $f$  in our bands of interest. We also show that for any continuous-time multiband signal that is also approximately time-limited, the resulting finite-length sample vector can be well-approximated by the multiband modulated DPSS dictionary. This result serves as a supplement to [13], which shows this approximation guarantee is available for a time-limited signal which has its spectrum concentrated in the bands of interest.

We hope that these results will prove useful in the continued study and application of multiband modulated DPSS dictionaries.

The remainder of this paper is organized as follows. Section 2 defines the time- and multiband-limiting operator and provides some important background information on DPSS’s. We state our main results in Section 3. We conclude in Section 4 with a final discussion.

## 2 Preliminaries

### 2.1 Notation

Finite-dimensional vectors and matrices are indicated by bold characters. We index all such vectors and matrices beginning at 0. The Hermitian transpose of a matrix  $\mathbf{A}$  is denoted by  $\mathbf{A}^H$ . For any natural number  $N$ , we let  $[N]$  denote the set  $\{0, 1, \dots, N - 1\}$ . For any  $k \in \{1, 2, \dots, N\}$ , let  $[\mathbf{A}]_k$  denote the  $N \times k$  matrix formed by taking the first  $k$  columns of  $\mathbf{A} \in \mathbb{C}^{N \times N}$ . 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  $N \rightarrow \infty$ .

<sup>1</sup>By equivalent, we mean that  $\mathbf{B}_{N, \mathbb{W}} \mathbf{x} = \mathcal{I}_N (\mathcal{B}_{\mathbb{W}} (\mathcal{I}_N^* (\mathbf{x})))$  for any  $\mathbf{x} \in \mathbb{C}^N$ .

<sup>2</sup>By “bands of interest,” we mean the union of intervals  $\mathbb{F}$  for continuous-time signals and  $\mathbb{W}$  for discrete-time signals. We assume these bands are known and are used to construct the multiband modulated DPSS dictionary. The results in this paper, however, can also have application in the problem of detecting the active bands from a set of possible candidates, as was studied in [13].

## 2.2 Time, index, and multiband-limiting operators

To begin, let  $\mathcal{B}_{\mathbb{W}} : \ell_2(\mathbb{Z}) \rightarrow \ell_2(\mathbb{Z})$  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

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

where  $\star$  stands for convolution. In addition, let  $\mathcal{T}_N : \ell_2(\mathbb{Z}) \rightarrow \ell_2(\mathbb{Z})$  denote the operator that zeros out all entries outside the index range  $\{0, 1, \dots, N-1\}$ . 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  $\mathcal{I}_N : \ell_2(\mathbb{Z}) \rightarrow \mathbb{C}^N$  as

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

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

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

We can observe that  $\mathcal{T}_N = \mathcal{I}_N^* \mathcal{I}_N$ .

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

$$\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y))[m] := \sum_{n=0}^{N-1} \left( y[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), \quad m \in \mathbb{Z}. \quad (8)$$

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

$$\mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y)))[m] = \begin{cases} \sum_{n=0}^{N-1} \left( y[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), & m \in [N], \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

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

$$\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\mathbf{y}))) [m] = \sum_{n=0}^{N-1} \left( \mathbf{y}[n] \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \right), \quad m \in [N]. \quad (10)$$

Suppose  $y' \in \ell_2(\mathbb{Z})$  is an eigenfunction of  $\mathcal{T}_N \mathcal{B}_{\mathbb{W}} \mathcal{T}_N$  with corresponding eigenvalue  $\lambda'$ :  $\mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y')))) = \lambda' y'$ . 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  $\mathbf{y}''$  and  $\lambda''$  satisfy  $\mathcal{I}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{I}_N^*(\mathbf{y}''))) = \lambda'' \mathbf{y}''$ , then we can conclude that  $\mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(\mathcal{I}_N^*(\mathbf{y}'')))) = \lambda'' \mathcal{I}_N^*(\mathbf{y}'')$ . 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  $\mathcal{T}_N \mathcal{B}_{\mathbb{W}} \mathcal{T}_N$ .

Note that  $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$  is equivalent to the covariance matrix  $\mathbf{B}_{N, \mathbb{W}}$  (see (7)), as a linear operator on  $\mathbb{C}^N$ . Thus, in order to answer the questions raised in Section 1.3, we will study the eigenvalue concentration behavior of  $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ .

## 2.3 DPSS bases for sampled bandlimited signals

In this subsection, we briefly review important definitions and properties of DPSS's from [13, 37].

### 2.3.1 DPSS's and DPSS vectors

**Definition 2.1.** (DPSS's [37]) Given  $W \in (0, \frac{1}{2})$  and  $N \in \mathbb{N}$ , the Discrete Prolate Spheroidal Sequences (DPSS's)  $\{s_{N,W}^{(0)}, s_{N,W}^{(1)}, \dots, s_{N,W}^{(N-1)}\}$  are real-valued discrete-time sequences that satisfy  $\mathcal{B}_{[-W,W]}(\mathcal{T}_N(s_{N,W}^{(l)})) = \lambda_{N,W}^{(l)} s_{N,W}^{(l)}$  for all  $l \in [N]$ . Here  $\lambda_{N,W}^{(0)}, \dots, \lambda_{N,W}^{(N-1)}$  are the eigenvalues of the operator  $\mathcal{B}_{[-W,W]} \mathcal{T}_N$  with order  $1 > \lambda_{N,W}^{(0)} > \lambda_{N,W}^{(1)} > \dots > \lambda_{N,W}^{(N-1)} > 0$ .

The DPSS's are orthogonal on  $\mathbb{Z}$  and on  $\{0, 1, \dots, N-1\}$ , and they are normalized so that

$$\langle \mathcal{T}_N(s_{N,W}^{(k)}), \mathcal{T}_N(s_{N,W}^{(l)}) \rangle = \begin{cases} 1, & k = l \\ 0, & k \neq l. \end{cases}$$

Consequently,  $\|s_{N,W}^{(l)}\|_2^2 = (\lambda_{N,W}^{(l)})^{-1}$ . The vector obtained by index-limiting  $s_{N,W}^{(l)}$  to the index range  $\{0, 1, \dots, N-1\}$  is an eigenvector of the  $N \times N$  matrix  $\mathbf{B}_{N,W}$  with elements given by<sup>3</sup>

$$\mathbf{B}_{N,W}[m, n] := \int_{-W}^W e^{j2\pi f(m-n)} df = \frac{\sin(2\pi W(m-n))}{\pi(m-n)}.$$

DPSS's are useful for constructing a dictionary that efficiently represents index-limited versions of sampled bandlimited signals. As pointed out in [13], the index-limited DPSS's also satisfy  $\mathcal{I}_N(\mathcal{B}_{[-W,W]}(\mathcal{T}_N(s_{N,W}^{(l)}))) = \lambda_{N,W}^{(l)} \mathcal{I}_N(s_{N,W}^{(l)})$ .

**Definition 2.2.** (DPSS vectors [37]) Given  $W \in (0, \frac{1}{2})$  and  $N \in \mathbb{N}$ , the DPSS vectors  $\mathbf{s}_{N,W}^{(0)}, \mathbf{s}_{N,W}^{(1)}, \dots, \mathbf{s}_{N,W}^{(N-1)} \in \mathbb{R}^N$  are defined by index-limiting the DPSS's to the index range  $\{0, 1, \dots, N-1\}$ :

$$\mathbf{s}_{N,W}^{(l)} = \mathcal{I}_N(s_{N,W}^{(l)})$$

and satisfy

$$\mathcal{I}_N(\mathcal{B}_{[-W,W]}(\mathcal{I}_N^*(\mathbf{s}_{N,W}^{(l)}))) = \mathbf{B}_{N,W} \mathbf{s}_{N,W}^{(l)} = \lambda_{N,W}^{(l)} \mathbf{s}_{N,W}^{(l)}.$$

It follows that  $\mathbf{B}_{N,W}$  can be factorized as

$$\mathbf{B}_{N,W} = \mathbf{S}_{N,W} \mathbf{\Lambda}_{N,W} \mathbf{S}_{N,W}^H,$$

where  $\mathbf{\Lambda}_{N,W}$  is an  $N \times N$  diagonal matrix whose diagonal elements are the DPSS eigenvalues  $\lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \dots, \lambda_{N,W}^{(N-1)}$  and  $\mathbf{S}_{N,W}$  is a square ( $N \times N$ ) matrix whose  $l$ -th column is the DPSS vector  $\mathbf{s}_{N,W}^{(l)}$  for all  $l \in [N]$ .

The following provides a useful result on the clustering of the eigenvalues  $\lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \dots, \lambda_{N,W}^{(N-1)}$ .

**Lemma 2.3.** (Clustering of eigenvalues [13, 37]) 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  $W, \epsilon$ ) and an integer  $N_0(W, \epsilon)$  (which may also depend on  $W, \epsilon$ ) such that

$$1 - \lambda_{N,W}^{(l)} \leq C_1(W, \epsilon) e^{-C_2(W, \epsilon)N}, \quad \forall l \leq \lfloor 2NW(1 - \epsilon) \rfloor \quad (11)$$

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  $W, \epsilon$ ) and an integer  $N_1(W, \epsilon)$  (which may also depend on  $W, \epsilon$ ) such that

$$\lambda_{N,W}^{(l)} \leq C_3(W, \epsilon) e^{-C_4(W, \epsilon)N}, \quad \forall l \geq \lceil 2NW(1 + \epsilon) \rceil \quad (12)$$

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

In words, the first  $\approx 2NW$  eigenvalues tend to cluster very close to 1, while the remaining eigenvalues tend to cluster very close to 0. As a consequence of this behavior, the effective dimensionality of the vectors  $\mathcal{M}_{[-W,W]} := \{\mathbf{e}_f\}_{f \in [-W,W]}$  (which trace out a finite-length curve in  $\mathbb{C}^N$ ) is essentially  $2NW$ , and we can use a subspace formed by the first  $\approx 2NW$  DPSS vectors to approximate this curve with low MSE.

<sup>3</sup>For convenience, we use  $\mathbf{B}_{N,W}$  instead of  $\mathbf{B}_{N,[-W,W]}$  to denote the matrix which is equivalent to the operator  $\mathcal{I}_N \mathcal{B}_{[-W,W]} \mathcal{I}_N^*$ . This is also the reason that we use  $\lambda_{N,W}, s_{N,W}$  and  $\mathbf{s}_{N,W}$  (which will be defined later) instead of  $\lambda_{N,[-W,W]}, s_{N,[-W,W]}$  and  $\mathbf{s}_{N,[-W,W]}$ .

### 2.3.2 DPSS bases for sampled bandpass signals

Let us now consider the eigenvectors of the operator  $\mathcal{I}_N(\mathcal{B}_{[f_c-W, f_c+W]}(\mathcal{I}_N^*))$ , which can be expressed as:

$$\begin{aligned} \mathcal{I}_N(\mathcal{B}_{[f_c-W, f_c+W]}(\mathcal{I}_N^*(\mathbf{y}))) [m] &= \left( \int_{f_c-W}^{f_c+W} e^{j2\pi m f} df \right) \star (\mathcal{I}_N^*(\mathbf{y})) [m] \\ &= \sum_{n=0}^{N-1} e^{j2\pi f_c(m-n)} \frac{\sin(2\pi W(m-n))}{\pi(m-n)} \mathbf{y}[n] \end{aligned}$$

for all  $m = 0, 1, \dots, N-1$ . Let  $\mathbf{E}_{f_c}$  denote an  $N \times N$  diagonal matrix with entries

$$\mathbf{E}_{f_c}[m, n] := \begin{cases} e^{j2\pi f_c m}, & m = n, \\ 0, & m \neq n. \end{cases}$$

We can verify that the modulated DPSS vectors  $\mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(0)}, \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(1)}, \dots, \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(N-1)}$  satisfy

$$\mathcal{I}_N(\mathcal{B}_{[f_c-W, f_c+W]}(\mathcal{I}_N^*(\mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(l)}))) = \mathbf{E}_{f_c} \mathbf{B}_{N,W} \mathbf{E}_{f_c}^H \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(l)} = \lambda_{N,W}^{(l)} \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(l)}.$$

That is,  $(\lambda_{N,W}^{(l)}, \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(l)})$  is an *eigenpair* of the operator  $\mathcal{I}_N(\mathcal{B}_{[f_c-W, f_c+W]}(\mathcal{I}_N^*))$  for all  $l \in [N]$ .

For any integer  $k \in \{1, 2, \dots, N\}$ , let  $\mathbf{Q} := [\mathbf{E}_{f_c} \mathbf{s}_{N,W}]_k$  denote the  $N \times k$  matrix formed by taking the first  $k$  modulated DPSS vectors. Also let  $\mathbf{P}_\mathbf{Q}$  denote the orthogonal projection onto the column space of  $\mathbf{Q}$ . It is shown in [13] that the dictionary  $\mathbf{Q}$  provides very accurate approximations (in an MSE sense) for finite-length sample vectors arising from sampling random bandpass signals.

**Theorem 2.4.** ([13] 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  $\mathbf{x} = [x(0) \ x(T_s) \ \dots \ x((N-1)T_s)]^T \in \mathbb{C}^N$  denote a finite vector of samples acquired from  $x(t)$  with a sampling interval of  $T_s \leq 1/(2 \max\{|F_c \pm \frac{B_{\text{band}}}{2}|\})$ . Let  $f_c = F_c T_s$  and  $W = \frac{B_{\text{band}} T_s}{2}$ . We will have

$$\mathbb{E} [\|\mathbf{x} - \mathbf{P}_\mathbf{Q} \mathbf{x}\|_2^2] = \frac{1}{2W} \int_{f_c-W}^{f_c+W} \|\mathbf{e}_f - \mathbf{P}_\mathbf{Q} \mathbf{e}_f\|_2^2 df = \frac{1}{2W} \sum_{l=k}^{N-1} \lambda_{N,W}^{(l)}.$$

Furthermore, for fixed  $\epsilon \in (0, \frac{1}{2W} - 1)$ , set  $k = 2NW(1 + \epsilon)$ . Then

$$\mathbb{E} [\|\mathbf{x} - \mathbf{P}_\mathbf{Q} \mathbf{x}\|_2^2] \leq \frac{C_3(W, \epsilon)}{2W} N e^{-C_4(W, \epsilon)N} \quad (13)$$

for all  $N \geq N_1(W, \epsilon)$ , where  $N_1(W, \epsilon)$ ,  $C_3(W, \epsilon)$ ,  $C_4(W, \epsilon)$  are constants specified in Lemma 2.3. For comparison,  $\mathbb{E} [\|\mathbf{x}\|_2^2] = \|\mathbf{e}_f\|_2^2 = N$ .

## 3 Main Results

We now consider the multiband case, where  $\mathbb{W} \subseteq [-\frac{1}{2}, \frac{1}{2}]$  is a union of  $J$  intervals as in (5). For each  $i \in [J]$ , define  $\Psi_i = [\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}]_{k_i}$  for some value  $k_i \in \{1, 2, \dots, N\}$  that we can choose as desired. We construct the multiband modulated DPSS dictionary  $\Psi$  by concatenating these subdictionaries:

$$\Psi := [\Psi_0 \ \Psi_1 \ \dots \ \Psi_{J-1}]. \quad (14)$$

In this section, we investigate the efficiency of using  $\Psi$  to represent discrete-time sinusoids and sampled multiband signals.



### 3.1 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  $\mathbf{B}_{N,\mathbb{W}}$ ), which reveals the effective dimensionality of the finite union of curves  $\mathcal{M}_\mathbb{W} = \{\mathbf{e}_f\}_{f \in \mathbb{W}}$ .

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.

**Lemma 3.1.** *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

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

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

**Proof.** See Appendix A.

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.

**Theorem 3.2.** *Suppose  $\mathbb{W}$  is a finite union of  $J$  pairwise disjoint intervals as defined in (5). 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  $\varepsilon$  and  $1 - \varepsilon$  is bounded above by*

$$\#\{l : \varepsilon \leq \lambda_{N,\mathbb{W}}^{(l)} \leq 1 - \varepsilon\} \leq J \frac{\frac{2}{\pi^2} \log(N-1) + \frac{2}{\pi^2} \frac{2N-1}{N-1}}{\varepsilon(1-\varepsilon)}. \quad (15)$$

**Proof.** See Appendix B.

This result states that the number of eigenvalues in  $[\varepsilon, 1 - \varepsilon]$  is in the order of  $\log(N)$  for any fixed  $\varepsilon \in (0, \frac{1}{2})$ . 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  $N|\mathbb{W}| = \sum_i 2NW_i$ .

**Theorem 3.3.** *Let  $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$  be a finite union of  $J$  disjoint intervals having the form in (5). 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*

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

**Proof.** See Appendix C.

Note that results similar to the above two theorems for time-frequency localization in the continuous domain have been established in [22, 25, 28]. Similar to the ideas used in [22], the key to proving Theorem 3.2 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 3.3. The proof techniques of [25, 28] form the basis of our analysis in Appendix C, 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 2.3), 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. This is captured formally in the following result.

**Theorem 3.4.** Let  $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$  be a fixed finite union of  $J$  disjoint intervals having the form in (5).

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

$$\lambda_{N,\mathbb{W}}^{(l)} \geq 1 - \bar{C}_1(\mathbb{W}, \epsilon)N^2 e^{-\bar{C}_2(\mathbb{W}, \epsilon)N}, \quad \forall l \leq J - 1 + \sum_i [2NW_i(1 - \epsilon)]$$

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

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

$$\lambda_{N,\mathbb{W}}^{(l)} \leq \bar{C}_3(\mathbb{W}, \epsilon)e^{-\bar{C}_4(\mathbb{W}, \epsilon)N}, \quad \forall l \geq \sum_i [2NW_i(1 + \epsilon)]$$

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

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

**Proof.** See Appendix D.

## 3.2 Multiband modulated DPSS dictionaries for sampled multiband signals

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

$$\Phi := [\mathbf{u}_{N,\mathbb{W}}^{(0)} \mathbf{u}_{N,\mathbb{W}}^{(1)} \cdots \mathbf{u}_{N,\mathbb{W}}^{(p-1)}], \quad (16)$$

where  $\mathbf{u}_{N,\mathbb{W}}^{(l)}, \forall l \in [N]$  are the eigenvectors of  $\mathcal{L}_N \mathbf{B}_{\mathbb{W}} \mathcal{L}_N^*$ . Let  $\Psi$  be the multiband modulated DPSS dictionary defined in (14).

There are three main reasons why the dictionary  $\Psi$  may be useful representing sampled multiband signals. First, direct computation of  $\Phi$  is difficult due to the clustering of the eigenvalues of  $\mathbf{B}_{N,\mathbb{W}}$ . However, in the single band case, the matrix  $\mathbf{B}_{N,W}$  is known to commute with a symmetric tridiagonal matrix that has well-separated eigenvalues, and hence its eigenvectors can be efficiently and stably computed [37]. Grünbaum [21] gave a certain condition for a Toeplitz matrix to commute with a tridiagonal matrix with a simple spectrum. We can check that the matrix  $\mathbf{B}_{N,\mathbb{W}}$  in general does not satisfy this condition, except for the case when  $\mathbb{W}$  consists of only a single interval. However, we emphasize that  $\Psi$  is constructed simply by modulating DPSS's, which, again, can be computed efficiently.

Second, the multiband modulated DPSS dictionary  $\Psi$  provides an efficient representation for sampled multiband signals. Davenport and Wakin [13] 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 3.2.3. 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 3.2.2.

Third, as indicated by the results in Section 3.1,  $\approx N|\mathbb{W}|$  dictionary atoms are necessary in order to achieve a high degree of approximation for the discrete-time sinusoids in an MSE sense. Our results, along with [13], show that the multiband modulated DPSS dictionary  $\Psi$  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  $\Psi$  and by the columns of the optimal dictionary  $\Phi$ .

### 3.2.1 The subspace angle between $\mathcal{S}_\Psi$ and $\mathcal{S}_\Phi$

In order to compare subspaces of possibly different dimensions, we require the following definition of angle between subspaces.

**Definition 3.5.** Let  $\mathcal{S}_\Psi$  and  $\mathcal{S}_\Phi$  be the subspaces formed by the columns of the matrices  $\Psi$  and  $\Phi$  respectively. The subspace angle  $\Theta_{\mathcal{S}_\Psi, \mathcal{S}_\Phi}$  between  $\mathcal{S}_\Psi$  and  $\mathcal{S}_\Phi$  is given by

$$\cos(\Theta_{\mathcal{S}_\Psi, \mathcal{S}_\Phi}) := \inf_{\phi \in \mathcal{S}_\Phi, \|\phi\|_2=1} \|\mathbf{P}_\Psi \phi\|_2$$

if  $\dim(\mathcal{S}_\Psi) \geq \dim(\mathcal{S}_\Phi)$ , or

$$\cos(\Theta_{\mathcal{S}_\Psi, \mathcal{S}_\Phi}) := \inf_{\psi \in \mathcal{S}_\Psi, \|\psi\|_2=1} \|\mathbf{P}_\Phi \psi\|_2$$

if  $\dim(\mathcal{S}_\Psi) < \dim(\mathcal{S}_\Phi)$ . Here  $\mathbf{P}_\Psi$  (or  $\mathbf{P}_\Phi$ ) denotes the orthogonal projection onto the column space of  $\Psi$  (or  $\Phi$ ).

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

**Theorem 3.6.** Let  $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$  be a fixed finite union of  $J$  disjoint intervals having the form in (5). Fix  $\epsilon \in (0, \min\{1, \frac{1}{|\mathbb{W}|} - 1\})$ . Let  $p = \sum_i \lceil 2NW_i(1 + \epsilon) \rceil$  and  $\Phi$  be the  $N \times p$  matrix defined in (16). Also let  $k_i \leq \lfloor 2NW_i(1 - \epsilon) \rfloor, \forall i \in [J]$  and  $\Psi$  be the matrix defined in (14). Then for any column  $\psi$  in  $\Psi$ ,

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

and

$$\cos(\Theta_{\mathcal{S}_\Psi, \mathcal{S}_\Phi}) \geq \sqrt{\frac{1 - \kappa_1(N, \mathbb{W}, \epsilon) - N\sqrt{\kappa_1(N, \mathbb{W}, \epsilon)} - 3N\sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)}e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N}}{1 + 3N\sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)}e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N}}} \quad (17)$$

if  $N \geq \max\{\bar{N}_0(\mathbb{W}, \epsilon), \bar{N}_1(\mathbb{W}, \epsilon)\}$ . Here  $\tilde{C}_1(\mathbb{W}, \epsilon) = \max\{C_1(W_i, \epsilon), \forall i \in [J]\}$ ,  $\tilde{C}_2(\mathbb{W}, \epsilon) = \min\{C_2(W_i, \epsilon), \forall i \in [J]\}$ ,  $\bar{N}_0(\mathbb{W}, \epsilon)$ ,  $\bar{N}_1(\mathbb{W}, \epsilon)$ ,  $\tilde{C}_3(\mathbb{W}, \epsilon)$ , and  $\tilde{C}_4(\mathbb{W}, \epsilon)$  are the constants specified in Theorem 3.4, and  $C_1(W_i, \epsilon)$  and  $C_2(W_i, \epsilon)$  are the constants specified in Lemma 2.3.

**Proof.** See Appendix E.

We can also guarantee that the subspace angle between  $\mathcal{S}_\Psi$  and  $\mathcal{S}_\Phi$  is small if, in constructing  $\Psi$ , each  $k_i$  is chosen slightly larger than  $2NW_i$ , and in constructing  $\Phi$ , we take  $p$  to be slightly smaller than  $\sum_i 2NW_i$ . This result is established in Corollary 3.8, which follows from Theorem 3.7.

**Theorem 3.7.** Let  $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$  be a finite union of  $J$  disjoint intervals having the form in (5). Given some values  $k_i \in \{1, 2, \dots, N\}, \forall i \in [J]$ , let  $\Psi$  be the matrix defined in (14). Then

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

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

**Proof.** See Appendix F.

**Corollary 3.8.** Let  $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$  be a fixed finite union of  $J$  disjoint intervals having the form in (5). 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  $\Phi$  be the  $N \times p$  matrix defined in (16). Also let  $k_i = \lceil 2NW_i(1 + \epsilon) \rceil, \forall i \in [J]$  and  $\Psi$  be the matrix defined in (14). Then for any column  $\mathbf{u}_{N, \mathbb{W}}^{(l)}$  in  $\Phi$ ,

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

and

$$\cos(\Theta_{\mathcal{S}_\Psi, \mathcal{S}_\Phi}) \geq \sqrt{1 - 2\kappa_2(N, \mathbb{W}, \epsilon) + \kappa_2^2(N, \mathbb{W}, \epsilon) - N\sqrt{2\kappa_2(N, \mathbb{W}, \epsilon) - \kappa_2^2(N, \mathbb{W}, \epsilon)}} \quad (18)$$

for all  $N \geq \max\{\bar{N}_0(\mathbb{W}, \epsilon), \bar{N}_1(\mathbb{W}, \epsilon)\}$ , where  $\bar{N}_0(\mathbb{W}, \epsilon)$ ,  $\bar{N}_1(\mathbb{W}, \epsilon)$ ,  $\tilde{C}_1(\mathbb{W}, \epsilon)$ ,  $\tilde{C}_2(\mathbb{W}, \epsilon)$ ,  $\tilde{C}_3(\mathbb{W}, \epsilon)$  and  $\tilde{C}_4(\mathbb{W}, \epsilon)$  are constants specified in Theorem 3.4, and  $\kappa_2(N, \mathbb{W}, \epsilon)$  is defined as  $\kappa_2(N, \mathbb{W}, \epsilon) := \tilde{C}_1(\mathbb{W}, \epsilon)N^2 e^{-\tilde{C}_2(\mathbb{W}, \epsilon)N} + N\tilde{C}_3(\mathbb{W}, \epsilon)e^{-\tilde{C}_4(\mathbb{W}, \epsilon)N}$ .

**Proof.** See Appendix G.

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

### 3.2.2 Approximation quality for discrete-time sinusoids

The above results show that  $\Psi$  spans nearly the same space as  $\Phi$  in the case where both dictionaries contain  $\approx N|\mathbb{W}|$  columns. In this section, we investigate the approximation quality of  $\Psi$  for discrete-time sinusoids with frequencies in the bands of interest. Then, in the next section, we investigate the approximation quality of  $\Psi$  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 [37]. We review this result in Appendix H.

**Theorem 3.9.** *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  $\epsilon'$ ) such that*

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

for all  $N \geq N_0(W', \epsilon')$ , where  $N_0(W', \epsilon')$  and  $C_2(W', \epsilon')$  are constants defined in Lemma 2.3.

**Proof.** See Appendix I.

To the best of our knowledge, this is the first work that rigorously shows that *every* discrete-time sinusoid with a frequency  $f \in [-W, W]$  is well-approximated by a DPSS basis  $[\mathcal{S}_{N,W}]_k$  with  $k$  slightly larger than  $2NW$ . This result extends the approximation guarantee in an MSE sense presented in [13]. We now extend this result for the multiband modulated DPSS dictionary.

**Corollary 3.10.** *Let  $\mathbb{W} \subset [-\frac{1}{2}, \frac{1}{2}]$  be a fixed finite union of  $J$  disjoint intervals having the form in (5). Fix  $\epsilon \in (0, \frac{1}{|\mathbb{W}|} - 1)$ . Let  $k_i = 2NW_i(1 + \epsilon), \forall i \in [J]$  and  $\Psi$  be the matrix defined in (14). Then there exist constants  $C_{10}(\mathbb{W}, \epsilon)$  and  $C_{11}(\mathbb{W}, \epsilon)$  (which may depend on  $\mathbb{W}$  and  $\epsilon$ ) and an integer  $N_2(\mathbb{W}, \epsilon)$  (which may also depend on  $\mathbb{W}$  and  $\epsilon$ ) such that*

$$\|e_f - \mathbf{P}_\Psi e_f\|_2^2 \leq C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N}, \quad \forall f \in \mathbb{W} \quad (19)$$

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

**Proof.** See Appendix J.

### 3.2.3 Approximation quality for sampled multiband signals (statistical analysis)

As indicated in [13], 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 [13, Theorem 4.4] to sampled multiband signals where each band has a possibly different width.

**Theorem 3.11.** *Suppose for each  $i \in [J]$ ,  $x_i(t)$  is a continuous-time, zero-mean, wide sense stationary random process with power spectrum*

$$P_{x_i}(F) = \begin{cases} \frac{1}{\sum_{i=0}^{J-1} B_{\text{band}_i}}, & F_i - \frac{B_{\text{band}_i}}{2} \leq F \leq F_i + \frac{B_{\text{band}_i}}{2}, \\ 0, & \text{otherwise,} \end{cases} \quad (20)$$

and furthermore suppose  $x_0(t), x_1(t), \dots, x_{J-1}(t)$  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  $\mathbf{x}_i = [x_i(0) \ x_i(T_s) \ \dots \ x_i((N-1)T_s)]^T \in \mathbb{C}^N$  denote a finite vector of samples acquired from  $x_i(t)$  and let  $\mathbf{x} = \sum_{i=1}^J \mathbf{x}_i$ . Set  $f_i = F_i T_s$  and  $W_i = \frac{B_{\text{band}_i} T_s}{2}$ . Let  $\Psi$  be the matrix defined in (14) for some given  $k_i$ . Then

$$\mathbb{E}[\|\mathbf{x} - \mathbf{P}_\Psi \mathbf{x}\|_2^2] \leq \frac{1}{|\mathbb{W}|} \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \lambda_{N, W_i}^{(l_i)}, \quad (21)$$

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

**Proof.** See Appendix K.

The right hand side of (21) can be made small by choosing  $k_i \approx 2NW_i$  for each  $i \in [J]$ ; recall Lemma 2.3. Aside from allowing for different band widths, the above result improves the upper bound of [13, 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 3.12.** *Suppose  $x$  is a continuous-time signal with Fourier transform  $X(F)$  supported on  $\mathbb{F} = \bigcup_{i=0}^{J-1} [F_i - B_{\text{band}_i}/2, F_i + B_{\text{band}_i}/2]$ , i.e.,*

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

*Let  $\mathbf{x} = [x(0) \ x(T_s) \ \dots \ x((N-1)T_s)]^T \in \mathbb{C}^N$  denote a finite vector of samples acquired from  $x(t)$  with a sampling interval of  $T_s \leq 1/(2 \max\{F_c \pm \frac{B_{\text{band}}}{2}\})$ . Let  $W_i = T_s B_{\text{band}_i}/2$ ,  $f_i = T_s F_i$  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  $\Psi$  be the matrix defined in (14). Then*

$$\|\mathbf{x} - \mathbf{P}_{\Psi} \mathbf{x}\|_2^2 \leq \left( \int_{\mathbb{W}} |\tilde{x}(f)|^2 df \right) \cdot C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N} \quad (22)$$

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 3.10.

**Proof.** See Appendix L.

Corollary 3.12 can be applied in various settings:

- The sequence  $x[n]$  encountered in most practical problems has finite energy. For example, if we assume that  $\int_{\mathbb{W}} |\tilde{x}(f)|^2 df \leq 1$ , we conclude that  $\|\mathbf{x} - \mathbf{P}_{\Psi} \mathbf{x}\|_2^2 \leq |\mathbb{W}| C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N}$ .
- Moreover, in some practical problems, the finite-energy sequence  $x[n]$  may be approximately time-limited to the index range  $n = 0, 1, \dots, N-1$  such that for some  $\delta$ ,  $\|\mathbf{x}\|_2^2 = \|\mathcal{I}_N(x)\|_2^2 \geq (1-\delta)\|\mathbf{x}\|_2^2$ . In this case, (22) guarantees that

$$\frac{\|\mathbf{x} - \mathbf{P}_{\Psi} \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \frac{\int_{\mathbb{W}} |\tilde{x}(f)|^2 df}{\|\mathbf{x}\|_2^2} \cdot C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N} \leq \frac{1}{1-\delta} C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N}, \quad (23)$$

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

Along with the result proved in [13] 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

$$\frac{\|\mathbf{s}_{N,W}^{(l)} - \mathbf{P}_{[\mathbf{s}_{N,W}]_k} \mathbf{s}_{N,W}^{(l)}\|_2}{\|\mathbf{s}_{N,W}^{(l)}\|_2} = 1 \quad (24)$$

even when we choose  $k = 2NW(1 + \epsilon)$ . In this case, the approximation guarantee in (24) is much worse than what appears in (23). Such examples are pathological, however: the infinite sequence  $\mathbf{s}_{N,W}^{(l)}$  has energy  $\|\mathbf{s}_{N,W}^{(l)}\|_2^2 = (\lambda_{N,W}^{(l)})^{-1}$ , which according to Lemma 2.3 is exponentially large when  $l \geq 2NW(1 + \epsilon)$ , and yet the energy of the sampled vector  $\|\mathbf{s}_{N,W}^{(l)}\|_2^2$  is only 1. Moreover, the spectrum of the infinite sequence  $\mathbf{s}_{N,W}^{(l)}$  is entirely concentrated in the band  $[-W, W]$  while the spectrum of the time-limited sequence  $\mathcal{T}_N(\mathbf{s}_{N,W}^{(l)})$  is almost entirely contained outside the band  $[-W, W]$ , and so on. Based on probabilistic guarantees such as Theorem 3.11, we conclude that such pathological examples are indeed relatively uncommon in practice. We refer to [13] for additional discussion of this topic.

## 4 Conclusions

In this paper, we have provided a thorough analysis of the spectrum of a time- and multiband-limiting operator in the discrete-time domain. We have showed that the information level of finite-length multiband sample vectors is essentially equal to their time-frequency area, which also indicates the number of dictionary atoms required in order to obtain a high-quality approximation. We have also considered the angle between the subspaces spanned by the eigenfunctions of the time- and multiband-limiting operator and by the multiband modulated DPSS dictionary. Our results show that the multiband modulated DPSS dictionary is nearly optimal in terms of representing finite-length vectors arising from sampling multiband analog signals.

We have showed that the multiband modulated DPSS dictionary can not only guarantee a very high degree of approximation accuracy in an MSE sense for finite-length multiband sample vectors, but also that it can guarantee such accuracy uniformly over all discrete-time sinusoids in the bands of interest. Though we are not guaranteed such accuracy uniformly over all sampled multiband signals, we have suggested that such accuracy holds for most practical problems involving multiband signals. Thus, our work supports the growing evidence that multiband modulated DPSS dictionaries can be useful for engineering applications.

## Acknowledgements

We gratefully acknowledge Mark Davenport, Armin Eftekhari, and Justin Romberg for valuable discussions and insightful comments.

## A Proof of Lemma 3.1

**Proof.** Let  $\mathbf{y} \in \mathbb{C}^N$ ,  $\mathbf{y} \neq \mathbf{0}$  be an arbitrary vector. Then

$$\begin{aligned} \langle \mathcal{I}_N(\mathcal{B}_W(\mathcal{I}_N^*(\mathbf{y}))), \mathbf{y} \rangle &= \sum_{m=0}^{N-1} \mathcal{I}_N(\mathcal{B}_W(\mathcal{I}_N^*(\mathbf{y}))) [m] \bar{\mathbf{y}}[m] = \sum_{m=0}^{N-1} \left( \sum_{n=0}^{N-1} \int_{\mathbb{W}} e^{j2\pi f(m-n)} df \mathbf{y}[n] \right) \bar{\mathbf{y}}[m] \\ &= \int_{\mathbb{W}} \left( \sum_{m=0}^{N-1} e^{j2\pi f m} \bar{\mathbf{y}}[m] \right) \left( \sum_{n=0}^{N-1} e^{-j2\pi f n} \mathbf{y}[n] \right) df = \int_{\mathbb{W}} \left| \sum_{n=0}^{N-1} \mathbf{y}[n] e^{-j2\pi f n} \right|^2 df > 0, \end{aligned}$$

where  $\bar{\mathbf{y}}$  is the complex-conjugate of the vector  $\mathbf{y}$ ,  $\sum_{n=0}^{N-1} \mathbf{y}[n] e^{-j2\pi f n}$  is the DTFT of  $\mathcal{I}_N^*(\mathbf{y})$ , and the last inequality is derived from the fact that compactly supported signals cannot have perfectly flat magnitude response.

By Parseval's Theorem, we know  $\int_{-1/2}^{1/2} \left| \sum_{n=0}^{N-1} \mathbf{y}[n] e^{-j2\pi f n} \right|^2 df = \|\mathbf{y}\|_2^2$ . Therefore

$$\langle \mathcal{I}_N(\mathcal{B}_W(\mathcal{I}_N^*(\mathbf{y}))), \mathbf{y} \rangle = \int_{\mathbb{W}} \left| \sum_{n=0}^{N-1} \mathbf{y}[n] e^{-j2\pi f n} \right|^2 df < \|\mathbf{y}\|_2^2.$$

Thus, we have

$$0 < \min_{\mathbf{y} \in \mathbb{C}^N} \frac{\langle \mathcal{I}_N(\mathcal{B}_W(\mathcal{I}_N^*(\mathbf{y}))), \mathbf{y} \rangle}{\|\mathbf{y}\|_2^2} \leq \lambda_{N,W}^{(l)} \leq \max_{\mathbf{y} \in \mathbb{C}^N} \frac{\langle \mathcal{I}_N(\mathcal{B}_W(\mathcal{I}_N^*(\mathbf{y}))), \mathbf{y} \rangle}{\|\mathbf{y}\|_2^2} < 1$$

for all  $l \in [N]$ .

By noting that  $\mathcal{I}_N \mathcal{B}_W \mathcal{I}_N^*$  is equivalent to  $\mathbf{B}_{N,W}$ , we have

$$\sum_{l=0}^{N-1} \lambda_{N,W}^{(l)} = \text{trace}(\mathbf{B}_{N,W}) = \sum_{n=0}^{N-1} \mathbf{B}_{N,W}[n, n] = \sum_{n=0}^{N-1} \int_{\mathbb{W}} e^{j2\pi f 0} df = N|\mathbb{W}|. \quad \square$$

## B Proof of Theorem 3.2

**Proof.** First we state a useful inequality about the Frobenius norm of positive semi-definite matrices. Suppose  $\mathbf{X} \in \mathbb{C}^{N \times N}$  and  $\mathbf{Y} \in \mathbb{C}^{N \times N}$  are two arbitrary positive semi-definite matrices. Then

$$\begin{aligned} \|\mathbf{X} + \mathbf{Y}\|_F^2 &= \text{trace}\left((\mathbf{X} + \mathbf{Y})^H(\mathbf{X} + \mathbf{Y})\right) \\ &= \|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2 + 2\text{trace}(\mathbf{X}^H\mathbf{Y}) \\ &\geq \|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2, \end{aligned}$$

where the last inequality is derived from the fact that  $\text{trace}(\mathbf{X}^H\mathbf{Y})$  is nonnegative, which can be showed as follows. By the hypothesis that  $\mathbf{X}$  and  $\mathbf{Y}$  are positive semi-definite matrices, we have the factorization  $\mathbf{X}^H = \mathbf{X} = \mathbf{X}^{1/2}\mathbf{X}^{1/2}$ , where  $\mathbf{X}^{1/2}$  is also a positive semi-definite matrix.<sup>4</sup> Then we conclude that  $\text{trace}(\mathbf{X}^H\mathbf{Y}) = \text{trace}(\mathbf{X}^{1/2}\mathbf{X}^{1/2}\mathbf{Y}) = \text{trace}(\mathbf{X}^{1/2}\mathbf{Y}\mathbf{X}^{1/2}) \geq 0$ , since  $\mathbf{X}^{1/2}\mathbf{Y}\mathbf{X}^{1/2}$  is also a positive semi-definite matrix.

We next bound the Frobenius norm of  $\mathbf{B}_{N,W_i}$  by

$$\begin{aligned} \|\mathbf{B}_{N,W_i}\|_F^2 &= N(2W_i)^2 + \sum_{m \neq n} \sum \left( \frac{\sin(2\pi W_i(m-n))}{\pi(m-n)} \right)^2 \\ &= 4NW_i^2 + 2 \sum_{n=1}^{N-1} (N-n) \left( \frac{\sin(2\pi W_i n)}{\pi n} \right)^2 \\ &= 4NW_i^2 + 2N \sum_{n=1}^{N-1} \left( \frac{\sin(2\pi W_i n)}{\pi n} \right)^2 - 2 \sum_{n=1}^{N-1} n \left( \frac{\sin(2\pi W_i n)}{\pi n} \right)^2 \\ &= 4NW_i^2 + 2N \left( W_i - 2W_i^2 - \sum_{n=N}^{\infty} \left( \frac{\sin(2\pi W_i n)}{\pi n} \right)^2 \right) - 2 \sum_{n=1}^{N-1} n \left( \frac{\sin(2\pi W_i n)}{\pi n} \right)^2 \\ &\geq 4NW_i^2 + 2N \left( W_i - 2W_i^2 - \frac{1}{\pi^2} \int_{N-1}^{\infty} \frac{1}{x^2} dx \right) - 2 \frac{1}{\pi^2} \left( \int_1^{N-1} \frac{1}{x} dx + 1 \right) \\ &= 2NW_i - \frac{2}{\pi^2} \frac{2N-1}{N-1} - \frac{2}{\pi^2} \log(N-1), \end{aligned}$$

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

$$\begin{aligned} \|\mathbf{B}_{N,\mathbb{W}}\|_F^2 &= \left\| \sum_{i=0}^{J-1} \mathbf{E}_{f_i} \mathbf{B}_{N,W_i} \mathbf{E}_{f_i}^H \right\|_F^2 \\ &\geq \sum_{i=0}^{J-1} \|\mathbf{B}_{N,W_i}\|_F^2 \\ &\geq \sum_{i=0}^{J-1} \left( 2NW_i - \frac{2}{\pi^2} \frac{2N-1}{N-1} - \frac{2}{\pi^2} \log(N-1) \right) \\ &= N|\mathbb{W}| - J \left( \frac{2}{\pi^2} \frac{2N-1}{N-1} + \frac{2}{\pi^2} \log(N-1) \right), \end{aligned}$$

where the second line follows since  $\mathbf{E}_{f_i} \mathbf{B}_{N,W_i} \mathbf{E}_{f_i}^H$  is positive semi-definite. Recalling the result stated in Lemma 3.1 that  $\sum_{l=0}^{N-1} \lambda_{N,\mathbb{W}}^{(l)} = \text{trace}(\mathbf{B}_{N,\mathbb{W}}) = N|\mathbb{W}|$ , we get

$$\sum_{l=0}^{N-1} \lambda_{N,\mathbb{W}}^{(l)} (1 - \lambda_{N,\mathbb{W}}^{(l)}) = \text{trace}(\mathbf{B}_{N,\mathbb{W}}) - \|\mathbf{B}_{N,\mathbb{W}}\|_F^2 \leq J \left( \frac{2}{\pi^2} \frac{2N-1}{N-1} + \frac{2}{\pi^2} \log(N-1) \right).$$

<sup>4</sup>Note that  $\mathbf{X}$  has the eigen-decomposition  $\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{V}^H$  where  $\mathbf{V}$  is an orthonormal matrix and  $\mathbf{D}$  is a diagonal matrix whose diagonal elements are non-negative, giving the square root  $\mathbf{X}^{1/2} = \mathbf{V}\mathbf{D}^{1/2}\mathbf{V}^H$ .

Thus, equation (15) follows by noting that for any  $\varepsilon \in (0, \frac{1}{2})$  one has

$$\sum_{l=0}^{N-1} \lambda_{N,\mathbb{W}}^{(l)} (1 - \lambda_{N,\mathbb{W}}^{(l)}) \geq \sum_{\{l: \varepsilon \leq \lambda_{N,\mathbb{W}}^{(l)} \leq 1 - \varepsilon\}} \lambda_{N,\mathbb{W}}^{(l)} (1 - \lambda_{N,\mathbb{W}}^{(l)}) \geq \varepsilon (1 - \varepsilon) \#\{l: \varepsilon \leq \lambda_{N,\mathbb{W}}^{(l)} \leq 1 - \varepsilon\}. \quad \square$$

## C Proof of Theorem 3.3

**Proof.** A precise proof of a similar result for time- and band-limiting operators in the continuous domain was first given in [28]. Izu and Lakey [25] 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 3.3. For convenience, we also use  $\lambda_{N,\mathbb{W}}^{(0)}, \lambda_{N,\mathbb{W}}^{(1)}, \dots, \lambda_{N,\mathbb{W}}^{(N-1)}$  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  $l \in [N]$ , the Weyl-Courant minimax representation of the eigenvalues can be stated as

$$\begin{aligned} \lambda_{N,\mathbb{W}}^{(l)} &= \begin{cases} \min_{S_l} \max_{y \in \ell_2(\mathbb{Z}), y \perp S_l} \frac{\langle \mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y))), y \rangle}{\langle y, y \rangle}, \\ \max_{S_{l+1}} \min_{y \in \ell_2(\mathbb{Z}), y \in S_{l+1}} \frac{\langle \mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y))), y \rangle}{\langle y, y \rangle}, \end{cases} \\ &= \begin{cases} \min_{S_l} \max_{y \in S([N]), y \perp S_l} \frac{\langle \mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y))), y \rangle}{\langle y, y \rangle}, \\ \max_{S_{l+1}} \min_{y \in S([N]), y \in S_{l+1}} \frac{\langle \mathcal{T}_N(\mathcal{B}_{\mathbb{W}}(\mathcal{T}_N(y))), y \rangle}{\langle y, y \rangle}, \end{cases} \\ &= \begin{cases} \min_{S_l} \max_{y \in S([N]), y \perp S_l} \frac{\int_{\mathbb{W}} |\tilde{y}(f)|^2 df}{\|y\|_2^2}, \\ \max_{S_{l+1}} \min_{y \in S([N]), y \in S_{l+1}} \frac{\int_{\mathbb{W}} |\tilde{y}(f)|^2 df}{\|y\|_2^2}, \end{cases} \end{aligned} \quad (25)$$

where  $S_l$  is an  $l$ -dimensional subspace of  $\ell_2(\mathbb{Z})$ , and  $\tilde{y}(f)$  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

$$\tilde{g}(f) = \begin{cases} \sqrt{2N} \cos(N\pi f) e^{-j2\pi f \lfloor \frac{N}{2} \rfloor}, & |f| \leq \frac{1}{2N}, \\ 0, & \frac{1}{2N} < |f| \leq \frac{1}{2}. \end{cases} \quad (26)$$

Then  $\|g\|_2^2 = 1$  and  $g[n] \geq \frac{1}{\sqrt{2N}}$  for all  $n \in [N]$ .

**Proof** (of Lemma C.1). First it is easy to check that  $\|g\|_2^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} |\tilde{g}(f)|^2 df = 1$ . Then computing the inverse DTFT directly yields

$$g[n] = \frac{1}{\sqrt{2N}} \operatorname{sinc}\left(\frac{n - \lfloor \frac{N}{2} \rfloor}{N} - \frac{1}{2}\right) + \frac{1}{\sqrt{2N}} \operatorname{sinc}\left(\frac{n - \lfloor \frac{N}{2} \rfloor}{N} + \frac{1}{2}\right).$$

Let  $\xi(t) = \operatorname{sinc}(t - \frac{1}{2}) + \operatorname{sinc}(t + \frac{1}{2})$ . Taking the derivative 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,  $g[n] \geq \frac{1}{\sqrt{2N}}$  since  $|\frac{n - \lfloor \frac{N}{2} \rfloor}{N}| \leq \frac{1}{2}$  for all  $n \in [N]$ .  $\square$

### C.1 Upper bound

From equation (25), we know that

$$\lambda_{N,\mathbb{W}}^{(l)} = \min_{S_l} \max_{y \in S([N]), y \perp S_l} \frac{\int_{\mathbb{W}} |\tilde{y}(f)|^2 df}{\|y\|_2^2}.$$

Therefore, in order to bound the eigenvalues from above, it suffices to pick an appropriate  $l$ -dimensional subspace  $S_l \subset \ell_2(\mathbb{Z})$  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 (26). Let  $\mathcal{E}_{f_0} : \ell_2(\mathbb{Z}) \rightarrow \ell_2(\mathbb{Z})$  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  $S_{\iota_+}$  be the  $\iota_+$ -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  $S_{\iota_+}$ , then

$$0 = \langle y, \mathcal{E}_{\frac{n'}{N}} g \rangle = \langle \tilde{y}, \tilde{g}(\cdot - \frac{n'}{N}) \rangle = (\tilde{y} \star \tilde{\bar{g}}) \left( \frac{n'}{N} \right) =: g_y[n'], \quad n' \in L_+,$$

where  $\bar{g} := g^*$  is the complex-conjugate of the sequence  $g$  and  $\tilde{\bar{g}}$  is the DTFT of  $\bar{g}$ .

Now it follows that

$$\begin{aligned} \sum_{n' = -\lfloor \frac{N}{2} \rfloor}^{\lfloor \frac{N-1}{2} \rfloor} |g_y[n']|^2 &= \sum_{n' \in L_+^C} |g_y[n']|^2 \\ &= \sum_{n' \in L_+^C} \left| \int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} \tilde{y}(f) \tilde{\bar{g}}(\frac{n'}{N} - f) df \right|^2 \\ &\leq \sum_{n' \in L_+^C} \left( \|g\|_2^2 \int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} |\tilde{y}(f)|^2 df \right) \\ &\leq \int_{f \notin \mathbb{W}} |\tilde{y}(f)|^2 df \\ &= \|y\|_2^2 - \int_{f \in \mathbb{W}} |\tilde{y}(f)|^2 df, \end{aligned} \tag{27}$$

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  $y \odot \bar{g}$  denote the pointwise product between  $y$  and  $\bar{g}$ , that is  $(y \odot \bar{g})[n] = y[n] \bar{g}[n]$ . Note that  $y \odot \bar{g}$  has the same support in time as  $y$ , namely  $[N]$ , and  $\{\frac{1}{\sqrt{N}} e^{\frac{j2\pi n'}{N}}, -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor\}$  forms an orthobasis (normalized DFT basis) for  $\mathbb{C}^N$ . We can rewrite  $g_y[n'] = e^{\frac{j2\pi n'}{N}} (y \odot \bar{g})$ , which can be viewed as the DFT of  $y \odot \bar{g}$ . Therefore, using Parseval's theorem, we acquire

$$\sum_{n' = -\lfloor \frac{N}{2} \rfloor}^{\lfloor \frac{N-1}{2} \rfloor} |g_y[n']|^2 = N \|y \odot \bar{g}\|_2^2 \geq \frac{1}{2} \|y\|_2^2$$

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 (27), we observe that

$$\frac{1}{2} \|y\|_2^2 \leq \|y\|_2^2 - \int_{f \in \mathbb{W}} |\tilde{y}(f)|^2 df,$$

and therefore,

$$\lambda_{N, \mathbb{W}}^{(\iota_+)} \leq \frac{\int_{\mathbb{W}} |\tilde{y}(f)|^2 df}{\|y\|_2^2} \leq \frac{1}{2}.$$

## C.2 Lower bound

In the other direction, consider the minimax representation

$$\lambda_{N, \mathbb{W}}^{(\iota)} = \max_{S_{\iota+1}} \min_{y \in S([N]), y \in S_{\iota+1}} \frac{\int_{\mathbb{W}} |\tilde{y}(f)|^2 df}{\|y\|_2^2}.$$

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 (26), let the time-limited sequence  $h \in \ell_2([N])$  be such that  $h[n] = 1/\bar{g}[n]$  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  $S_{\iota_-}$  be the  $\iota_-$ -dimensional subspace of  $\ell_2(\mathbb{Z})$  spanned by the functions  $\mathcal{E}_{\frac{n'}{N}} h, n' \in L_-$ , that is,

$$S_{\iota_-} := \text{span} \left( \{ \mathcal{E}_{\frac{n'}{N}} h \}_{n' \in L_-} \right).$$

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,

$$y \odot \bar{g} = \sum_{n' \in L_-} b_{n'} e_{\frac{n'}{N}}.$$

Noting that  $\{ \frac{1}{\sqrt{N}} e_{\frac{n'}{N}}, -\lfloor \frac{N}{2} \rfloor \leq n' \leq \lfloor \frac{N-1}{2} \rfloor \}$  forms an orthobasis for  $\mathbb{C}^N$ , we obtain

$$\sum_{n' \in L_-} |b_{n'}|^2 = N \|y \odot \bar{g}\|_2^2 = N \sum_{n=0}^{N-1} |y[n] \odot \bar{g}[n]|^2 \geq \frac{1}{2} \sum_{n=0}^{N-1} |y[n]|^2 = \frac{1}{2} \|y\|_2^2$$

since by definition,  $g[n] \geq \frac{1}{\sqrt{2N}}$  for all  $n \in [N]$ . On the other hand,

$$b_{n'} = \sum_{n=0}^{N-1} \bar{g}[n] y[n] e^{-j2\pi \frac{n'}{N} n} = \langle y, \mathcal{E}_{\frac{n'}{N}} g \rangle.$$

Now using the same procedure as in (27), one has

$$\begin{aligned} \sum_{n' \in L_-} |b_{n'}|^2 &= \sum_{n' \in L_-} |\langle y, \mathcal{E}_{\frac{n'}{N}} g \rangle|^2 \\ &= \sum_{n' \in L_-} \left| \int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} \tilde{y}(f) \tilde{g}(\frac{n'}{N} - f) df \right|^2 \\ &\leq \sum_{n' \in L_-} \left( \|g\|_2^2 \int_{\frac{n'-1/2}{N}}^{\frac{n'+1/2}{N}} |\tilde{y}(f)|^2 df \right) \\ &\leq \int_{f \in \mathbb{W}} |\tilde{y}(f)|^2 df, \end{aligned}$$

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  $y \in S([N])$ ),

$$\frac{1}{2} \|y\|_2^2 \leq \int_{f \in \mathbb{W}} |\tilde{y}(f)|^2 df.$$

And hence

$$\lambda_{N, \mathbb{W}}^{(\iota_- - 1)} \geq \frac{\int_{f \in \mathbb{W}} |\tilde{y}(f)|^2 df}{\|y\|_2^2} \geq \frac{1}{2}. \quad \square$$

## D Proof of Theorem 3.4

### D.1 Proof of eigenvalues that cluster near zero

**Proof.** Since  $\mathbf{B}_{N,\mathbb{W}} = \sum_{i=0}^{J-1} \mathbf{E}_{f_i} \mathbf{B}_{N,W_i} \mathbf{E}_{f_i}^H$ , according to [24] (see pp. 181), the following holds

$$\lambda_{N,\mathbb{W}}^{(l)} \leq \sum_{i=0}^{J-1} \lambda_{N,W_i}^{(l_i)}$$

for all  $l_i \in [N], i \in [J]$  and  $l = \sum_{i=0}^{J-1} l_i \in [N]$ .

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 2.3 with respect to  $W_i$  and  $\epsilon$ . If we let  $\bar{N}_1(\mathbb{W}, \epsilon) = \max\{N_1(W_i, \epsilon), \forall i \in [J]\}$ , then we have

$$\lambda_{N,W_i}^{(l_i)} \leq C_3(W_i, \epsilon) e^{-C_4(W_i, \epsilon)N}, \forall l_i \geq \lceil 2NW_i(1 + \epsilon) \rceil, i \in [J]$$

for all  $N \geq \bar{N}_1(\mathbb{W}, \epsilon)$ . Hence, by choosing  $l_i \geq \lceil 2NW_i(1 + \epsilon) \rceil, \forall i \in [J]$ , we have

$$\lambda_{N,\mathbb{W}}^{(l)} \leq \sum_{i=0}^{J-1} C_3(W_i, \epsilon) e^{-C_4(W_i, \epsilon)N} \leq \bar{C}_3(\mathbb{W}, \epsilon) e^{-\bar{C}_4(\mathbb{W}, \epsilon)N},$$

for all  $N \geq \bar{N}_1(\mathbb{W}, \epsilon)$  and  $l \geq \sum_i \lceil 2NW_i(1 + \epsilon) \rceil$ , where  $\bar{C}_3(\mathbb{W}, \epsilon) = J \max\{C_3(W_i, \epsilon), \forall i \in [J]\}$  and  $\bar{C}_4(\mathbb{W}, \epsilon) = \min\{C_4(W_i, \epsilon), \forall i \in [J]\}$ .  $\square$

### D.2 $\epsilon$ -pseudo eigenvalue and eigenvectors

**Definition D.1.** ( $\epsilon$ -pseudo eigenvalue and eigenvector [34]) Let  $\mathbf{X} \in \mathbb{C}^{N \times N}$  be any matrix and  $\mathbf{u} \in \mathbb{C}^N$  be any vector with unit  $l_2$ -norm. Given  $\epsilon > 0$ , the number  $\lambda \in \mathbb{C}$  and vector  $\mathbf{u} \in \mathbb{C}^N$  are an  $\epsilon$ -pseudo eigenpair of  $\mathbf{X}$  if the following condition is satisfied:

$$\|(\mathbf{X} - \lambda \mathbf{I})\mathbf{u}\|_2^2 \leq \epsilon.$$

**Lemma D.2.** Suppose  $\mathbb{W}$  is a fixed finite union of  $J$  pairwise disjoint intervals as defined in (5). Fix  $\epsilon \in (0, 1)$ . For each  $i \in [J]$ , let  $N_0(W_i, \epsilon)$  be the constant specified in Lemma 2.3 with respect to  $W_i$  and  $\epsilon$  and let  $\tilde{N}_0(\mathbb{W}, \epsilon) = \max\{N_0(W_i, \epsilon), \forall i \in [J]\}$ . Then for all  $l_i \leq 2NW_i(1 - \epsilon), i \in [J]$  and  $N > \tilde{N}_0(\mathbb{W}, \epsilon)$ ,  $(\lambda_{N,W_i}^{(l_i)}, \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)})$  is an  $\epsilon$ -pseudo eigenpair of  $\mathcal{I}_N \mathbf{B}_{\mathbb{W}} \mathcal{I}_N^*$  with  $\epsilon \leq 2C_1(W_i, \epsilon) e^{-C_2(W_i, \epsilon)N}$ , or in detail

$$\mathcal{I}_N(\mathbf{B}_{\mathbb{W}}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)}))) = \lambda_{N,W_i}^{(l_i)} \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} + \mathbf{o}_i^{(l_i)},$$

where  $\mathbf{o}_i^{(l_i)} = \mathcal{I}_N(\mathbf{B}_{\mathbb{W} \setminus [f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)})))$  and  $\|\mathbf{o}_i^{(l_i)}\|_2^2 \leq 2C_1(W_i, \epsilon) e^{-C_2(W_i, \epsilon)N}$ . 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 2.3 corresponding to  $W_i$  and  $\epsilon$  for all  $i \in [J]$ .

**Proof** (of Lemma D.2). According to the definition of the operator  $\mathcal{I}_N \mathbf{B}_{\mathbb{W}} \mathcal{I}_N^*$ ,

$$\begin{aligned} & \left( \mathcal{I}_N(\mathbf{B}_{\mathbb{W}}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)}))) \right) [m] \\ &= \sum_{n=0}^{N-1} \sum_{i'=0}^{J-1} e^{j2\pi f_{i'}(m-n)} \frac{\sin(2\pi W_{i'}(m-n))}{\pi(m-n)} e^{j2\pi f_i n} \mathbf{s}_{N,W_i}^{(l_i)} [n] \\ &= e^{j2\pi f_i m} \lambda_{N,W_i}^{(l_i)} \mathbf{s}_{N,W_i}^{(l_i)} [m] + \sum_{n=0}^{N-1} \sum_{i'=0, i' \neq i}^{J-1} e^{j2\pi f_{i'}(m-n)} \frac{\sin(2\pi W_{i'}(m-n))}{\pi(m-n)} e^{j2\pi f_i n} \mathbf{s}_{N,W_i}^{(l_i)} [n] \\ &= e^{j2\pi f_i m} \lambda_{N,W_i}^{(l_i)} \mathbf{s}_{N,W_i}^{(l_i)} [m] + \mathcal{I}_N(\mathbf{B}_{\mathbb{W} \setminus [f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)}))) [m]. \end{aligned}$$

In what follows, we will bound the energy of  $\mathbf{o}_i^{(l_i)} = \mathcal{I}_N(\mathcal{B}_{\mathbb{W} \setminus [f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)})))$  as

$$\begin{aligned}
\|\mathbf{o}_i^{(l_i)}\|_2^2 &= \|\mathcal{I}_N(\mathcal{B}_{\mathbb{W} \setminus [f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)})))\|_2^2 \\
&\leq \|\mathcal{B}_{\mathbb{W} \setminus [f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)}))\|_2^2 \\
&\leq \|\mathcal{B}_{[-\frac{1}{2}, \frac{1}{2}] \setminus [f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)}))\|_2^2 \\
&= \|\mathbf{s}_{N, W_i}^{(l_i)}\|_2^2 - \|\mathcal{B}_{[f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)}))\|_2^2 \\
&\leq \|\mathbf{s}_{N, W_i}^{(l_i)}\|_2^2 - \|\mathcal{I}_N(\mathcal{B}_{[f_i - W_i, f_i + W_i]}(\mathcal{I}_N^*(\mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)})))\|_2^2 \\
&\leq 1 - (\lambda_{N, W_i}^{(l_i)})^2 \leq 1 - (1 - C_1(W_i, \epsilon))e^{-C_2(W_i, \epsilon)N} \\
&= 2C_1(W_i, \epsilon)e^{-C_2(W_i, \epsilon)N} - (C_1(W_i, \epsilon))e^{-C_2(W_i, \epsilon)N} \leq 2C_1(W_i, \epsilon)e^{-C_2(W_i, \epsilon)N}
\end{aligned}$$

for all  $l_i \leq \lfloor 2NW_i(1 - \epsilon) \rfloor$ ,  $i \in [J]$  and  $N \geq \tilde{N}_0(\mathbb{W}, \epsilon)$ . Here the second inequality in the sixth line follows simply from Lemma 2.3 since  $\tilde{N}_0(\mathbb{W}, \epsilon) \geq N_0(W_i, \epsilon)$ .  $\square$

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.

### D.3 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 D.3.** ([13] Lemma 5.1) Fix  $\epsilon \in (0, 1)$ . Let  $k_i = \lfloor 2NW_i(1 - \epsilon) \rfloor$ ,  $\forall i \in [J]$ , and let  $\Psi$  be the dictionary as defined in (14). Then for any pair of distinct columns  $\psi_1$  and  $\psi_2$  in  $\Psi$ , we have

$$|\langle \psi_1, \psi_2 \rangle| \leq 3\sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)} e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \quad (28)$$

and

$$\|\Psi^H \Psi\|_2 \leq 1 + 3N\sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)} e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N}$$

if  $N \geq \tilde{N}_0(\mathbb{W}, \epsilon)$ , where  $\tilde{C}_1(\mathbb{W}, \epsilon) = \max\{C_1(W_i, \epsilon), \forall i \in [J]\}$  and  $\tilde{C}_2(\mathbb{W}, \epsilon) = \min\{C_2(W_i, \epsilon), \forall i \in [J]\}$ . Here  $\|\Psi^H \Psi\|_2$  is the spectral norm (or largest singular value) of  $\Psi^H \Psi$ .

**Lemma D.4.** ([24]) Let  $\mathbf{X} \in \mathbb{C}^{N \times N}$  be a Hermitian matrix, and let  $\lambda_0(\mathbf{X}), \lambda_1(\mathbf{X}), \dots, \lambda_{N-1}(\mathbf{X})$  be its eigenvalues arranged in decreasing order. Then,

$$\lambda_0(\mathbf{X}) + \lambda_1(\mathbf{X}) + \dots + \lambda_{r-1}(\mathbf{X}) = \max_{\mathbf{U} \in \mathbb{C}^{N \times r}, \mathbf{U}^H \mathbf{U} = \mathbf{I}_r} \text{trace}(\mathbf{U}^H \mathbf{X} \mathbf{U}),$$

where  $\mathbf{I}_r$  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 D.5.** Let  $\mathbf{X} \in \mathbb{C}^{N \times N}$  be a positive-semidefinite (PSD) matrix, and let  $\lambda_0(\mathbf{X}), \lambda_1(\mathbf{X}), \dots, \lambda_{N-1}(\mathbf{X})$  be its eigenvalues arranged in decreasing order. Then, for any matrix  $\mathbf{M} \in \mathbb{C}^{N \times r}$ ,  $1 \leq r \leq N$ , the following inequality holds

$$\lambda_0(\mathbf{X}) + \lambda_1(\mathbf{X}) + \dots + \lambda_{r-1}(\mathbf{X}) \geq \text{trace}(\mathbf{M}^H \mathbf{X} \mathbf{M}) / \|\mathbf{M}^H \mathbf{M}\|_2.$$

**Proof** (of Lemma D.5). Let  $\sigma_0(\mathbf{M}), \dots, \sigma_{r-1}(\mathbf{M})$  denote the decreasing singular values of the matrix  $\mathbf{M}$ . Denote  $\mathbf{M} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^H$  as the truncated SVD of  $\mathbf{M}$ , where  $\mathbf{\Sigma}_r$  is an  $r \times r$  diagonal matrix with  $\sigma_0(\mathbf{M}), \dots, \sigma_{r-1}(\mathbf{M})$  along its diagonal.

Now applying Lemma D.4, we obtain

$$\begin{aligned}
\sum_{l=0}^{r-1} \lambda_l(\mathbf{X}) &\geq \text{trace}(\mathbf{U}_r^H \mathbf{X} \mathbf{U}_r) \\
&\geq \text{trace}(\mathbf{\Sigma}_r \mathbf{U}_r^H \mathbf{X} \mathbf{U}_r \mathbf{\Sigma}_r) / (\sigma_0(\mathbf{M}))^2 \\
&= \text{trace}(\mathbf{V}_r \mathbf{\Sigma}_r \mathbf{U}_r^H \mathbf{X} \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^H) / \|\mathbf{M}^H \mathbf{M}\|_2 \\
&= \text{trace}(\mathbf{M}^H \mathbf{X} \mathbf{M}) / \|\mathbf{M}^H \mathbf{M}\|_2,
\end{aligned}$$

where the first line follows directly from Lemma D.4, the second line is obtained because  $\mathbf{U}_r^H \mathbf{X} \mathbf{U}_r$  is PSD and hence its main diagonal elements are non-negative, and the third line follows because  $\mathbf{V}_r$  is an orthobasis and  $(\sigma_0(\mathbf{M}))^2 = \|\mathbf{M}^H \mathbf{M}\|_2$ .  $\square$

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  $\Psi$  be the dictionary as defined in (14). We have

$$\begin{aligned}
\sum_{l=0}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(l)} &\geq \text{trace} \left( \Psi^H \mathbf{B}_{N,\mathbb{W}} \Psi \right) / \left\| \Psi^H \Psi \right\|_2 \\
&= \left( \sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon) \rfloor} \left( \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} \right)^H \mathcal{I}_N \left( \mathcal{B}_{\mathbb{W}} \left( \mathcal{I}_N^* \left( \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} \right) \right) \right) \right) / \left\| \Psi^H \Psi \right\|_2 \\
&= \left( \sum_{i=0}^{J-1} \sum_{l_i=0}^{\lfloor 2NW_i(1-\epsilon) \rfloor} \left( \left( \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} \right)^H \left( \lambda_{N,W_i}^{(l_i)} \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} + \mathbf{o}_i^{(l_i)} \right) \right) \right) / \left\| \Psi^H \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}^{(l_i)} - \|\mathbf{o}_i^{(l_i)}\|_2 \right) \right) / \left\| \Psi^H \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{\tilde{C}_1(\mathbb{W}, \epsilon)} e^{-\frac{\tilde{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 - \tilde{C}_1(\mathbb{W}, \epsilon) e^{-\tilde{C}_2(\mathbb{W}, \epsilon)N} - \sqrt{2} \sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)} e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \right) \right)}{\left( 1 + 3N \sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)} e^{-\frac{\tilde{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{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N}}{1 + 3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{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{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \right) \left( 1 - 3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \right)}{\left( 1 + 3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)N}{2}} \right) \left( 1 - 3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \right)} \\
&\geq \frac{J + \sum_i \lfloor 2NW_i(1 - \epsilon) \rfloor - 6N^2 C_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} + \left( 3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \right)^2}{1 - \left( 3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N} \right)^2} \\
&\geq J + \sum_i \lfloor 2NW_i(1 - \epsilon) \rfloor - 6N^2 C_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}N}
\end{aligned}$$

for all  $N \geq \max\{\tilde{N}_0(\mathbb{W}, \epsilon), N'(\mathbb{W}, \epsilon)\}$ , where  $N'(\mathbb{W}, \epsilon) = \max\left\{\left(\frac{4}{C_2(\mathbb{W}, \epsilon)}\right)^2, \frac{4}{C_2(\mathbb{W}, \epsilon)} \log(3C_5(\mathbb{W}, \epsilon))\right\}$  is the constant such that  $3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)N}{2}} < 1$  for all  $N \geq N'(\mathbb{W}, \epsilon)$ .<sup>5</sup> Here the first line follows directly from Lemma D.5, the second line follows because  $\text{trace}(\Psi^H \mathbf{B}_{N,\mathbb{W}} \Psi) = \text{trace}\left(\sum_{i=0}^{J-1} \Psi_i^H \mathbf{B}_{N,\mathbb{W}} \Psi_i\right)$  and  $\mathbf{B}_{N,\mathbb{W}}$  is equivalent to  $\mathcal{I}_N \mathcal{B}_{\mathbb{W}} \mathcal{I}_N^*$ , the third line follows from Lemma D.2, the fourth line follows from the Cauchy-Schwarz inequality which indicates that  $|\left(\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)}\right)^H \mathbf{o}_i^{(l_i)}| \leq \|\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)}\|_2 \|\mathbf{o}_i^{(l_i)}\|_2 = \|\mathbf{o}_i^{(l_i)}\|_2$ , the fifth line follows from Lemmas 2.3, D.2 and D.3, the seventh line follows by setting  $C_5(\mathbb{W}, \epsilon) = \max\{\tilde{C}_1(\mathbb{W}, \epsilon), \sqrt{\tilde{C}_1(\mathbb{W}, \epsilon)}\}$ , the ninth line follows because  $J + \sum_i \lfloor 2NW(1 - \epsilon) \rfloor \leq N$ , and the last line follows because by assumption  $3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)N}{2}} < 1$ .

<sup>5</sup>This can be verified as  $3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)N}{2}} = 3C_5(\mathbb{W}, \epsilon) e^{-N\left(\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2} - \frac{\log N}{N}\right)} \leq 3C_5(\mathbb{W}, \epsilon) e^{-N\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{4}} \leq 1$  for all  $N \geq \max\left\{\left(\frac{4}{C_2(\mathbb{W}, \epsilon)}\right)^2, \frac{4}{C_2(\mathbb{W}, \epsilon)} \log(3C_5(\mathbb{W}, \epsilon))\right\}$ . 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  $N \geq \left(\frac{4}{C_2(\mathbb{W}, \epsilon)}\right)^2$ .

By noting that  $0 < \lambda_{N,\mathbb{W}}^{(N-1)} \leq \lambda_{N,\mathbb{W}}^{(0)} < 1$  from Lemma 3.1, we acquire

$$\begin{aligned} \lambda_{N,\mathbb{W}}^{(l)} &= \left( \sum_{l'=0}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(l')} \right) - \left( \sum_{l'=0, l' \neq l}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(l')} \right) \\ &\geq \left( \sum_{l'=0}^{J-1+\sum_i \lfloor 2NW_i(1-\epsilon) \rfloor} \lambda_{N,\mathbb{W}}^{(l')} \right) - \left( J-1 + \sum_i \lfloor 2NW_i(1-\epsilon) \rfloor \right) \\ &\geq 1 - 6N^2 C_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2} N} \end{aligned}$$

for all  $l \leq J-1 + \sum_i \lfloor 2NW_i(1-\epsilon) \rfloor$ , where the second line follows by setting  $\lambda_{N,\mathbb{W}}^{(l')}$ ,  $l' \neq l$  to 1. Fix  $\mathbb{W}$  and  $\epsilon$ . It is always possible to find a constant  $N'$  such that  $3NC_5(\mathbb{W}, \epsilon) e^{-\frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2} N} < 1$  for all  $N \geq N'$ . Now, for convenience, we set  $\bar{C}_1(\mathbb{W}, \epsilon) = 6C_5(\mathbb{W}, \epsilon)$ ,  $\bar{C}_2(\mathbb{W}, \epsilon) = \frac{\tilde{C}_2(\mathbb{W}, \epsilon)}{2}$ , and  $\bar{N}_0(\mathbb{W}, \epsilon) = \max\{\bar{N}_0(\mathbb{W}, \epsilon), N'\}$ . This completes the proof of Theorem 3.4.  $\square$

## E Proof of Theorem 3.6

**Proof.** First denote the eigen-decomposition of  $\mathbf{B}_{N,\mathbb{W}}$  as

$$\mathbf{B}_{N,\mathbb{W}} = \mathbf{U}_{N,\mathbb{W}} \mathbf{\Lambda}_{N,\mathbb{W}} \mathbf{U}_{N,\mathbb{W}}^H,$$

where  $\mathbf{\Lambda}_{N,\mathbb{W}}$  is an  $N \times N$  diagonal matrix whose diagonal elements are the eigenvalues  $\lambda_{N,\mathbb{W}}^{(0)}, \lambda_{N,\mathbb{W}}^{(1)}, \dots, \lambda_{N,\mathbb{W}}^{(N-1)}$  and  $\mathbf{U}_{N,\mathbb{W}}$  is a square ( $N \times N$ ) matrix defined by

$$\mathbf{U}_{N,\mathbb{W}} := [\mathbf{u}_{N,\mathbb{W}}^{(0)} \ \mathbf{u}_{N,\mathbb{W}}^{(1)} \ \dots \ \mathbf{u}_{N,\mathbb{W}}^{(N-1)}].$$

Also let  $\mathbf{a} = \mathbf{U}_{N,\mathbb{W}}^H \boldsymbol{\psi}$  be the coefficients of  $\boldsymbol{\psi}$  represented by  $\mathbf{U}_{N,\mathbb{W}}$ .

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 D.2, we have

$$\mathbf{B}_{N,\mathbb{W}} \boldsymbol{\psi} = \lambda_{N,W_i}^{(l_i)} \boldsymbol{\psi} + \mathbf{o}_i^{(l_i)}$$

for some  $l_i \leq \lfloor 2NW_i(1-\epsilon) \rfloor$ .

Plugging the eigen-decomposition of the matrix  $\mathbf{U}_{N,\mathbb{W}}$  into the above equation, we require

$$\mathbf{\Lambda}_{N,\mathbb{W}} \mathbf{a} = \lambda_{N,W_i}^{(l_i)} \mathbf{a} + \hat{\mathbf{o}}_i^{(l_i)},$$

where  $\hat{\mathbf{o}}_i^{(l_i)} = \mathbf{U}_{N,\mathbb{W}}^H \mathbf{o}_i^{(l_i)}$ . The elementary form of the above equation is

$$\lambda_{N,\mathbb{W}}^{(m)} \mathbf{a}[m] = \lambda_{N,W_i}^{(l_i)} \mathbf{a}[m] + \hat{\mathbf{o}}_i^{(l_i)}[m]$$

for all  $m \in [N]$ .

Now we have

$$\begin{aligned} \|\boldsymbol{\psi} - \mathbf{P}_{\Phi} \boldsymbol{\psi}\|_2^2 &= \sum_{m=\sum_i \lceil 2NW_i(1+\epsilon) \rceil}^{N-1} |\mathbf{a}[m]|^2 = \sum_{m=\sum_i \lceil 2NW_i(1+\epsilon) \rceil}^{N-1} \frac{|\hat{\mathbf{o}}_i^{(l_i)}[m]|^2}{\left| \lambda_{N,W_i}^{(l_i)} - \lambda_{N,\mathbb{W}}^{(m)} \right|^2} \\ &\leq \frac{\sum_{m=\sum_i \lceil 2NW_i(1+\epsilon) \rceil}^{N-1} |\hat{\mathbf{o}}_i^{(l_i)}[m]|^2}{\left( 1 - \tilde{C}_1(\mathbb{W}, \epsilon) e^{-\tilde{C}_2(\mathbb{W}, \epsilon) N} - \bar{C}_3(\mathbb{W}, \epsilon) e^{-\bar{C}_4(\mathbb{W}, \epsilon) N} \right)^2} \\ &\leq \frac{\|\mathbf{o}_i^{(l_i)}\|^2}{\left( 1 - \tilde{C}_1(\mathbb{W}, \epsilon) e^{-\tilde{C}_2(\mathbb{W}, \epsilon) N} - \bar{C}_3(\mathbb{W}, \epsilon) e^{-\bar{C}_4(\mathbb{W}, \epsilon) N} \right)^2} \\ &\leq \frac{2\tilde{C}_1(\mathbb{W}, \epsilon) e^{-\tilde{C}_2(\mathbb{W}, \epsilon) N}}{\left( 1 - \tilde{C}_1(\mathbb{W}, \epsilon) e^{-\tilde{C}_2(\mathbb{W}, \epsilon) N} - \bar{C}_3(\mathbb{W}, \epsilon) e^{-\bar{C}_4(\mathbb{W}, \epsilon) N} \right)^2} \end{aligned} \tag{29}$$

for all  $N \geq \max\{\bar{N}_0(\mathbb{W}, \epsilon), \bar{N}_1(\mathbb{W}, \epsilon)\}$ , where the second line follows by bounding the  $\lambda_{N,W_i}^{(l_i)}$  term using  $1 - C_1(W_i, \epsilon) e^{-C_2(W_i, \epsilon) N}$  (which is not less than  $1 - \tilde{C}_1(\mathbb{W}, \epsilon) e^{-\tilde{C}_2(\mathbb{W}, \epsilon) N}$ ) from Lemma 2.3 and bounding the

$\lambda_{N, \mathbb{W}}^{(m)}$  terms using Theorem 3.4, and the fourth line follows because  $\|\mathbf{o}_i^{(l_i)}\|^2 \leq 2C_1(W_i, \epsilon)e^{-C_2(W_i, \epsilon)N} \leq 2\tilde{C}_1(\mathbb{W}, \epsilon)e^{-\tilde{C}_2(\mathbb{W}, \epsilon)N}$ .

The following general result will help in extending (29) to an angle between the subspaces.

**Lemma E.1.** *Let  $\mathcal{S}_U$  and  $\mathcal{S}_V$  be the subspaces spanned by the columns of the matrices  $U \in \mathbb{C}^{N \times q}$  and  $V \in \mathbb{C}^{N \times r}$ , respectively. Here  $r \leq q \leq N$ . Suppose each column of  $V$  is normalized so that  $\|\mathbf{v}_l\|_2 = 1$  and is close to  $\mathcal{S}_U$  such that for some  $\delta_1$ ,  $\|\mathbf{v}_l - \mathbf{P}_U \mathbf{v}_l\|_2^2 \leq \delta_1$  for all  $l \in [r]$ . Furthermore, suppose the columns of  $V$  are approximately orthogonal to each other such that for some  $\delta_2$ ,  $|\langle \mathbf{v}_k, \mathbf{v}_l \rangle| \leq \delta_2$  for all  $k \neq l$ . Then we have*

$$\cos(\Theta_{\mathcal{S}_U \mathcal{S}_V}) \geq \sqrt{\frac{1 - \delta_1 - N(\delta_2 + \sqrt{\delta_1})}{1 + N\delta_2}}.$$

**Proof** (of Lemma E.1). Any  $\mathbf{v} \in \mathcal{S}_V$  can be written as a linear combination of  $\mathbf{v}_l$  in the form  $\mathbf{v} = \sum_l \alpha_l \mathbf{v}_l$ . We first bound the  $l_2$  norm of  $\mathbf{v}$  by

$$\begin{aligned} \|\mathbf{v}\|_2^2 &= \left\| \sum_{l=0}^{r-1} \alpha_l \mathbf{v}_l \right\|_2^2 \\ &= \sum_{l=0}^{r-1} |\alpha_l|^2 \|\mathbf{v}_l\|_2^2 + \sum_{l=0}^{r-1} \sum_{k=0, k \neq l}^{r-1} \langle \alpha_l \mathbf{v}_l, \alpha_k \mathbf{v}_k \rangle \\ &\leq \sum_{l=0}^{r-1} |\alpha_l|^2 + \sum_{l=0}^{r-1} \sum_{k=0, k \neq l}^{r-1} |\alpha_l| |\alpha_k| \delta_2 \\ &\leq \sum_{l=0}^{r-1} |\alpha_l|^2 + \sum_{l=0}^{r-1} \sum_{k=0, k \neq l}^{r-1} \frac{|\alpha_l|^2 + |\alpha_k|^2}{2} \delta_2 \\ &= \left( \sum_{l=0}^{r-1} |\alpha_l|^2 \right) (1 + (r-1)\delta_2) \leq \left( \sum_{l=0}^{r-1} |\alpha_l|^2 \right) (1 + N\delta_2), \end{aligned}$$

where the third line follows from the hypothesis that  $|\langle \mathbf{v}_k, \mathbf{v}_l \rangle| \leq \delta_2$  for all  $k \neq l$ . Similarly,

$$\begin{aligned} \|\mathbf{P}_U \mathbf{v}\|_2^2 &= \left\| \sum_{l=0}^{r-1} \mathbf{P}_U (\alpha_l \mathbf{v}_l) \right\|_2^2 \\ &= \sum_{l=0}^{r-1} |\alpha_l|^2 \|\mathbf{P}_U \mathbf{v}_l\|_2^2 + \sum_{l=0}^{r-1} \sum_{k=0, k \neq l}^{r-1} \langle \alpha_l \mathbf{P}_U \mathbf{v}_l, \alpha_k \mathbf{P}_U \mathbf{v}_k \rangle \\ &= \sum_{l=0}^{r-1} |\alpha_l|^2 \|\mathbf{P}_U \mathbf{v}_l\|_2^2 + \sum_{l=0}^{r-1} \sum_{k=0, k \neq l}^{r-1} \langle \alpha_l \mathbf{v}_l, \alpha_k (\mathbf{v}_k - (\mathbf{v}_k - \mathbf{P}_U \mathbf{v}_k)) \rangle \\ &\geq \sum_{l=0}^{r-1} |\alpha_l|^2 (1 - \delta_1) - \sum_{l=0}^{r-1} \sum_{k=0, k \neq l}^{r-1} |\alpha_l| |\alpha_k| (\delta_2 + \sqrt{\delta_1}) \\ &= \left( \sum_{l=0}^{r-1} |\alpha_l|^2 \right) (1 - \delta_1 - (r-1)(\delta_2 + \sqrt{\delta_1})) \geq \left( \sum_{l=0}^{r-1} |\alpha_l|^2 \right) (1 - \delta_1 - N(\delta_2 + \sqrt{\delta_1})), \end{aligned}$$

where the fourth line follows because  $\langle \mathbf{v}_l, \mathbf{v}_k - \mathbf{P}_U \mathbf{v}_k \rangle \leq \|\mathbf{v}_l\|_2 \|\mathbf{v}_k - \mathbf{P}_U \mathbf{v}_k\|_2 \leq \sqrt{\delta_1}$  and  $|\langle \mathbf{v}_k, \mathbf{v}_l \rangle| \leq \delta_2$  for all  $k \neq l$ .

Therefore, for any non-zero vector  $\mathbf{v} \in \mathcal{S}_V$  we have

$$\frac{\|\mathbf{P}_U \mathbf{v}\|_2^2}{\|\mathbf{v}\|_2^2} \geq \frac{1 - \delta_1 - N(\delta_2 + \sqrt{\delta_1})}{1 + N\delta_2}. \quad \square$$

Finally, (17) follows from Lemma E.1 by replacing  $U$  with  $\Phi$  and  $V$  with  $\Psi$ , and assigning  $\delta_1$  with the upper bound in (29) and  $\delta_2$  with the upper bound in (28).  $\square$

## F Proof of Theorem 3.7

**Proof.** For each  $i \in [J]$ , define  $\bar{\Psi}_i = [\mathbf{E}_{f_i} \mathbf{S}_{N, W_i} \sqrt{\Lambda_{N, W_i}}]_{k_i}$  for some given  $k_i \in \{1, 2, \dots, N\}$ . We construct the scaled multiband modulated DPSS matrix  $\bar{\Psi}$  by<sup>6</sup>

$$\bar{\Psi} := [\bar{\Psi}_0 \bar{\Psi}_1 \cdots \bar{\Psi}_{J-1}]. \quad (30)$$

The main idea is to bound  $\|\mathbf{P}_{\Psi} \mathbf{u}_{N, W}^{(l)}\|_2$  using  $\|\bar{\Psi} \bar{\Psi}^H \mathbf{u}_{N, W}^{(l)}\|_2$ . In order to use this argument, we first give out some useful results.

**Lemma F.1.** *Suppose  $\bar{\Psi}$  is the matrix defined in (30) with some given  $k_i \in \{1, 2, \dots, N\}, \forall i \in [J]$ . Then*

$$\|\bar{\Psi}\|_2 \leq 1.$$

**Proof** (of Lemma F.1) Let  $\mathbf{y} \in \mathbb{C}^N$ . Then

$$\begin{aligned} \|\bar{\Psi}^H \mathbf{y}\|_2^2 &= \sum_{i=0}^{J-1} \sum_{l_i=0}^{k_i-1} |\langle \mathbf{y}, \mathbf{E}_{f_i} \sqrt{\lambda_{N, W_i}^{(l_i)}} \mathbf{s}_{N, W_i}^{(l_i)} \rangle|^2 \\ &= \sum_{i=0}^{J-1} \sum_{l_i=0}^{k_i-1} \langle \mathbf{y}, \mathbf{E}_{f_i} \sqrt{\lambda_{N, W_i}^{(l_i)}} \mathbf{s}_{N, W_i}^{(l_i)} \rangle \langle \mathbf{E}_{f_i} \sqrt{\lambda_{N, W_i}^{(l_i)}} \mathbf{s}_{N, W_i}^{(l_i)}, \mathbf{y} \rangle \\ &= \sum_{i=0}^{J-1} \sum_{l_i=0}^{k_i-1} \mathbf{y}^H \mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)} \lambda_{N, W_i}^{(l_i)} (\mathbf{s}_{N, W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{y} \\ &\leq \sum_{i=0}^{J-1} \sum_{l_i=0}^{N-1} \mathbf{y}^H \mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)} \lambda_{N, W_i}^{(l_i)} (\mathbf{s}_{N, W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{y} \\ &= \sum_{i=0}^{J-1} \mathbf{y}^H \mathbf{E}_{f_i} \mathcal{I}_N(\mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y}))) = \sum_{i=0}^{J-1} \langle \mathcal{I}_N(\mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y}))), \mathbf{E}_{f_i}^H \mathbf{y} \rangle \\ &= \sum_{i=0}^{J-1} \langle \mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y})), \mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y}) \rangle = \sum_{i=0}^{J-1} \langle \mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y})), \mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y})) \rangle = \sum_{i=0}^{J-1} \|\mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{E}_{f_i}^H \mathbf{y}))\|_2^2 \\ &= \sum_{i=0}^{J-1} \int_{f_i - W_i}^{f_i + W_i} |\tilde{\mathbf{y}}(f)|^2 df = \int_{-1/2}^{1/2} \left( \sum_{i=0}^{J-1} \mathbb{1}_{[f_i - W_i, f_i + W_i)}(f) \right) |\tilde{\mathbf{y}}(f)|^2 df, \end{aligned}$$

where the fourth line follows because  $\mathbf{y}^H \mathbf{E}_{f_i} \mathbf{s}_{N, W_i}^{(l_i)} \lambda_{N, W_i}^{(l_i)} (\mathbf{s}_{N, W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{y} = \|\sqrt{\lambda_{N, W_i}^{(l_i)}} (\mathbf{s}_{N, W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{y}\|_2^2 \geq 0$ , the fifth line follows because  $\sum_{l_i=0}^{N-1} \mathbf{s}_{N, W_i}^{(l_i)} \lambda_{N, W_i}^{(l_i)} (\mathbf{s}_{N, W_i}^{(l_i)})^H \mathbf{x} = \mathcal{I}_N(\mathcal{B}_{W_i}(\mathcal{I}_N^*(\mathbf{x})))$ , and we use  $\tilde{\mathbf{y}}(f) = \sum_{n=0}^{N-1} \mathbf{y}[n] e^{-j2\pi f n}$  as the DTFT of  $\mathcal{I}_N^*(\mathbf{y})$  in the last three equations.

Noting that  $\sum_{i=0}^{J-1} \mathbb{1}_{[f_i - W_i, f_i + W_i)}(f) \leq 1$  for all  $f \in [-\frac{1}{2}, \frac{1}{2}]$  since we assume there is no overlap between each interval  $[f_i - W_i, f_i + W_i)$ , we conclude

$$\|\bar{\Psi}^H \mathbf{y}\|_2^2 \leq \int_{-1/2}^{1/2} |\tilde{\mathbf{y}}(f)|^2 df = \|\mathbf{y}\|_2^2$$

and

$$\|\bar{\Psi}\|_2 \leq 1. \quad \square$$

**Lemma F.2.** *For any  $k_i \in \{1, 2, \dots, N\}, i \in [J]$ , let  $\Psi$  and  $\bar{\Psi}$  be the matrices defined in (14) and (30) respectively. Then for any  $\mathbf{y} \in \mathbb{C}^{N \times 1}$ ,*

$$\|\mathbf{P}_{\Psi} \mathbf{y}\|_2 \geq \|\bar{\Psi} \bar{\Psi}^H \mathbf{y}\|_2. \quad (31)$$

<sup>6</sup>Hogan and Lakey [23] considered the scaled and shifted Prolate Spheroidal Wave Functions (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.



**Proof** (of Lemma F.2) Let  $\overline{\Psi} = \mathbf{U}_{\overline{\Psi}} \Sigma_{\overline{\Psi}} \mathbf{V}_{\overline{\Psi}}^H$  be a reduced SVD of  $\overline{\Psi}$ , where both  $\mathbf{U}_{\overline{\Psi}}$  and  $\mathbf{V}_{\overline{\Psi}}$  are orthonormal matrices of the proper dimension, and  $\Sigma_{\overline{\Psi}}$  is a diagonal matrix whose diagonal elements are the non-zero singular values of  $\overline{\Psi}$ . We have

$$\begin{aligned} \|\overline{\Psi} \overline{\Psi}^H \mathbf{y}\|_2 &= \|\mathbf{U}_{\overline{\Psi}} \Sigma_{\overline{\Psi}}^2 \mathbf{U}_{\overline{\Psi}}^H \mathbf{y}\|_2 \\ &\leq \|\mathbf{U}_{\overline{\Psi}}^H \mathbf{y}\|_2 \\ &= \|\mathbf{U}_{\overline{\Psi}} \mathbf{U}_{\overline{\Psi}}^H \mathbf{y}\|_2 \\ &= \|\mathbf{P}_{\Psi} \mathbf{y}\|_2 \end{aligned}$$

where the second line follows because  $\|\overline{\Psi}\|_2 \leq 1$  and hence the diagonal elements  $\Sigma_{\overline{\Psi}}$  are bounded above by 1, and the fourth line follows because each column in  $\overline{\Psi}$  is in also  $\Psi$  and hence  $\|\mathbf{P}_{\Psi} \mathbf{y}\|_2 = \|\mathbf{P}_{\overline{\Psi}} \mathbf{y}\|_2$ .  $\square$

Now we turn to prove Theorem 3.7. By (31), we observe that

$$\begin{aligned} \|\mathbf{P}_{\Psi} \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2 &\geq \|\overline{\Psi} \overline{\Psi}^H \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2 \\ &= \left\| \sum_{i=0}^{J-1} \sum_{l_i=k_i-1}^{k_i-1} \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} \lambda_{N,W_i}^{(l_i)} (\mathbf{s}_{N,W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{u}_{N,\mathbb{W}}^{(l)} \right\|_2 \\ &= \|\mathbf{B}_{N,\mathbb{W}} \mathbf{u}_{N,\mathbb{W}}^{(l)} - \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} \lambda_{N,W_i}^{(l_i)} (\mathbf{s}_{N,W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2 \\ &\geq \|\mathbf{B}_{N,\mathbb{W}} \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2 - \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \|\mathbf{E}_{f_i} \mathbf{s}_{N,W_i}^{(l_i)} \lambda_{N,W_i}^{(l_i)} (\mathbf{s}_{N,W_i}^{(l_i)})^H \mathbf{E}_{f_i}^H \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2 \\ &\geq \lambda_{N,\mathbb{W}}^{(l)} - \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \lambda_{N,W_i}^{(l_i)}. \quad \square \end{aligned}$$

## G Proof of Corollary 3.8

**Proof.** It follows from Theorem 3.7 that

$$\begin{aligned} \|\mathbf{P}_{\Psi} \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2 &\geq \lambda_{N,\mathbb{W}}^{(l)} - \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \lambda_{N,W_i}^{(l_i)} \\ &\geq 1 - \overline{C}_1(\mathbb{W}, \epsilon) N^2 e^{-\overline{C}_2(\mathbb{W}, \epsilon) N} - \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} C_3(W_i, \epsilon) e^{-C_4(W_i, \epsilon) N} \\ &\geq 1 - \overline{C}_1(\mathbb{W}, \epsilon) N^2 e^{-\overline{C}_2(\mathbb{W}, \epsilon) N} - \sum_{i=0}^{J-1} \sum_{l_i=k_i}^{N-1} \frac{1}{J} \overline{C}_3(\mathbb{W}, \epsilon) e^{-\overline{C}_4(\mathbb{W}, \epsilon) N} \\ &\geq 1 - \overline{C}_1(\mathbb{W}, \epsilon) N^2 e^{-\overline{C}_2(\mathbb{W}, \epsilon) N} - N \overline{C}_3(\mathbb{W}, \epsilon) e^{-\overline{C}_4(\mathbb{W}, \epsilon) N} \end{aligned}$$

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  $\lambda_{N,\mathbb{W}}^{(l)}$  term using Theorem 3.4 and by bounding the  $\lambda_{N,W_i}^{(l_i)}$  terms using Lemma 2.3, 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  $\|\mathbf{u}_{N,\mathbb{W}}^{(l)} - \mathbf{P}_{\Psi} \mathbf{u}_{N,\mathbb{W}}^{(l)}\|_2^2 \leq 2\kappa_2(N, \mathbb{W}, \epsilon) - \kappa_2^2(N, \mathbb{W}, \epsilon)$ . Noting also that  $\langle \mathbf{u}_{N,\mathbb{W}}^{(l)}, \mathbf{u}_{N,\mathbb{W}}^{(k)} \rangle = 0$  for all  $k \neq l$ , (18) follows directly from Lemma E.1.  $\square$

## H DTFT of DPSS vectors

The results presented in this appendix are useful in Appendix I, where we analyze the performance of the DPSS vectors for representing sampled pure tones inside the band of interest. Let  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f)$  denote the DTFT of the sequence  $\mathcal{T}_N(s_{N,W}^{(l)})$ , i.e.,  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f) = \sum_{n=0}^{N-1} s_{N,W}^{(l)}[n] e^{-j2\pi f n}$ . Figure 1 shows  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f)$

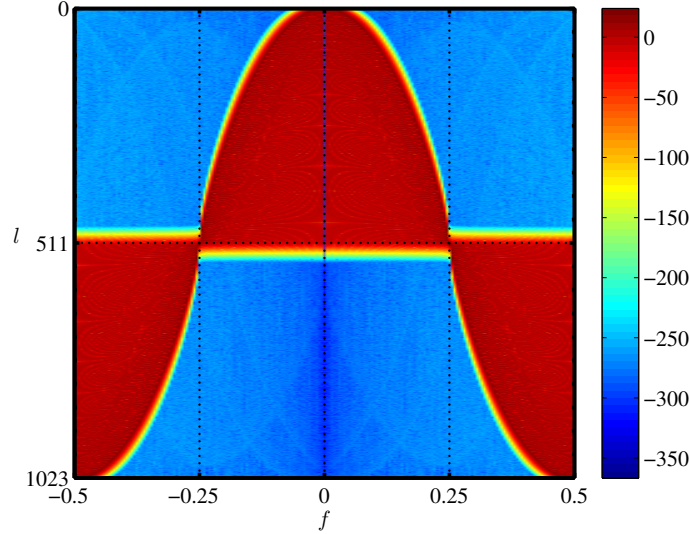


Figure 1: Illustration of  $\left|\tilde{\mathbf{s}}_{N,W}^{(l)}(f)\right|^2$ , or the energy in  $\{e_f\}$  captured by each DPSS vector. The horizontal axis stands for the digital frequency  $f$ , which ranges over  $[-\frac{1}{2}, \frac{1}{2}]$ , while the vertical axis stands for the index  $l \in [N]$ . The  $l$ -th horizontal line shows  $10 \log_{10} \left|\tilde{\mathbf{s}}_{N,W}^{(l)}(f)\right|^2$ . Here  $N = 1024$  and  $W = \frac{1}{4}$ .

for all  $l \in [N]$  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  $\lambda_{N,W}^{(l)}$  and  $\tilde{\mathbf{s}}_{N,W}(f)$  from [37].

**Lemma H.1.** ([37]) 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  $N \rightarrow \infty$ , we have

$$1 - \lambda_{N,W}^{(l)} \sim c_5^2 / \left(2\sqrt{2\alpha}\right)$$

and

$$\tilde{\mathbf{s}}_{N,W}^{(l)}(f) \sim \begin{cases} c_3 f_4(f), & W \leq |f| \leq \arccos(A - N^{-3/2})/2\pi, \\ c_5 f_5(f), & \arccos(A - N^{-3/2})/2\pi \leq |f| \leq 1/2. \end{cases}$$

Here

$$\begin{aligned} c_5 &= (l!)^{-1/2} \pi^{1/4} 2^{(14l+15)/8} \alpha^{(2l+3)/8} N^{(2l+1)/4} (\sqrt{2} + \sqrt{\alpha})^{-N} (2 - \alpha)^{(N-l-1/2)/2} \\ &= (l!)^{-1/2} \pi^{1/4} 2^{(14l+15)/8} \alpha^{(2l+3)/8} N^{(2l+1)/4} (2 - \alpha)^{-(l+1/2)/2} e^{-\frac{\gamma}{2}N}, \\ c_3 &= \pi^{1/2} 2^{-1/2} \alpha^{-1/4} [2 - \alpha]^{-1/4} N^{1/2} c_5 = O(N^{1/2}) c_5, \\ \gamma &= \log\left(1 + \frac{2\sqrt{\alpha}}{\sqrt{2} - \sqrt{\alpha}}\right), \\ f_4(f) &= J_0\left(\frac{N}{\sqrt{2} - \alpha} \sqrt{A - \cos(2\pi f)}\right), \\ f_5(f) &= \frac{\cos\left(\frac{N}{2} \arcsin(\theta(f)) + \frac{1}{2}(l + \frac{1}{2}) \arcsin(\phi(f)) + (l - N) \frac{\pi}{4} + \frac{3\pi}{8}\right)}{\left((A - \cos(2\pi f))(1 - \cos(2\pi f))\right)^{1/4}}, \\ \theta(f) &= \frac{\alpha + 2 \cos(2\pi f)}{2 - \alpha}, \quad \phi(f) = \frac{(2 - 3\alpha) - (2 + \alpha) \cos(2\pi f)}{(2 - \alpha)(1 - \cos(2\pi f))}, \end{aligned}$$

where  $J_0$  is the Bessel function of the first kind.

2. As  $N \rightarrow \infty$  and with  $l = \lfloor 2NW(1 - \epsilon') \rfloor$  for any  $\epsilon' \in (0, \epsilon]$ , we have

$$1 - \lambda_{N,W}^{(l)} \sim 2\pi L_2^{-1} d_6^2$$

and

$$\tilde{\mathbf{s}}_{N,W}^{(l)}(f) \sim \begin{cases} d_4 g_5(f), & W \leq |f| \leq \arccos(A - N^{-1})/2\pi, \\ d_6 g_6(f), & \arccos(A - N^{-1})/2\pi \leq |f| \leq 1/2. \end{cases}$$

Here

$$\begin{aligned} d_6 &= (L_2)^{-1/2} \pi^{1/2} 2^{1/2} e^{-CL_4/4} e^{-NL_3/2}, \\ d_4 &= (L_2)^{-1/2} \pi (1 - A^2)^{-1/4} e^{-CL_4/4} e^{-NL_3/2} N^{1/2}, \\ g_5(f) &= J_0 \left( N \sqrt{\frac{B - A}{1 - A^2}} (\cos(2\pi f) - A) \right), \\ g_6(f) &= R(f) \cos \left( \pi N \int_f^{1/2} \sqrt{\frac{B - \cos(2\pi t)}{A - \cos(2\pi t)}} dt + \frac{\pi C}{2} \int_f^{1/2} \frac{dt}{\sqrt{(B - \cos(2\pi t))(A - \cos(2\pi t))}} + \theta \right), \\ R(f) &= |(B - \cos(2\pi f))(A - \cos(2\pi f))|^{-1/4}, \quad C = \frac{1}{L_2} \pmod{\left( \frac{N}{2} L_1 + \left(2 + (-1)^l\right) \frac{\pi}{4}, 2\pi \right)}, \\ \theta &= \pmod{\left( \frac{\pi}{4} - \frac{N}{2} L_5 - \frac{C}{4} L_6, 2\pi \right)}, \\ L_1 &= \int_B^1 P(\xi) d\xi, \quad L_2 = \int_B^1 Q(\xi) d\xi, \quad L_3 = \int_A^B P(\xi) d\xi, \quad L_4 = \int_A^B Q(\xi) d\xi, \quad L_5 = \int_{-1}^A P(\xi) d\xi, \quad L_6 = L_2, \\ P(\xi) &= \left| \frac{\xi - B}{(\xi - A)(1 - \xi^2)} \right|^{1/2}, \quad Q(\xi) = |(\xi - B)(\xi - A)(1 - \xi^2)|^{-1/2}, \end{aligned}$$

where  $B$  is determined so that  $\int_B^1 \sqrt{\frac{\xi - B}{(\xi - A)(1 - \xi^2)}} d\xi = \frac{1}{N} \pi$  and  $\pmod{(y, 2\pi)}$  returns the remainder after division of  $y$  by  $2\pi$ .

## I Proof of Theorem 3.9

Noting that  $\mathbf{S}_{N,W}$  forms an orthobasis for  $\mathbb{C}^{N \times N}$ , the main idea is to show that the DPSS vectors  $\mathbf{s}_{N,W}^{(2NW(1+\epsilon))}, \mathbf{s}_{N,W}^{(2NW(1+\epsilon)+1)}, \dots, \mathbf{s}_{N,W}^{(N-1)}$  have their spectrum most concentrated outside of the band  $[-W, W]$ .

Since the sequence  $s_{N,W}^{(l)}$  is exactly bandlimited to the frequency range  $|f| \leq W$ , we know that its DTFT  $\tilde{s}_{N,W}^{(l)}(f) := \sum_{n=-\infty}^{\infty} s_{N,W}^{(l)}[n] e^{j2\pi f n}$  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  $n = 0, 1, \dots, N - 1$ , we may expect that  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f) := \sum_{n=0}^{N-1} \mathbf{s}_{N,W}^{(l)}[n] e^{j2\pi f n}$  is also approximately 0 for all  $W < |f| < \frac{1}{2}$  and  $l \leq 2NW(1 - \epsilon)$ . 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 [37] which states that  $|\tilde{\mathbf{s}}_{N,W}^{(l)}(f)| = |\tilde{\mathbf{s}}_{N, \frac{1}{2}-W}^{(N-1-l)}(\frac{1}{2}-f)|$ ,

we then have that the DPSS vectors  $\mathbf{s}_{N,W}^{(2NW(1+\epsilon))}, \mathbf{s}_{N,W}^{(2NW(1+\epsilon)+1)}, \dots, \mathbf{s}_{N,W}^{(N-1)}$  are almost orthogonal to any sinusoid with frequency inside the band  $[-W, W]$ .

Recall that  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f)$  is the DTFT of the sequence  $\mathcal{T}_N(s_{N,W}^{(l)})$ , i.e.,  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f) = \sum_{n=0}^{N-1} s_{N,W}^{(l)}[n] e^{-j2\pi f n}$ . We have

$$\langle \mathbf{s}_{N,W}^{(l)}, \mathbf{e}_f \rangle = \tilde{\mathbf{s}}_{N,W}^{(l)}(f),$$

for all  $l \in [N]$ . As we have observed in Figure 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 I.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  $\epsilon$ ) such that*

$$|\tilde{\mathbf{s}}_{N,W}^{(l)}(f)| \leq C_6(W, \epsilon) N^{3/4} e^{-\frac{C_2(W, \epsilon)}{2} N}, \quad W \leq |f| \leq 1/2$$

for all  $N \geq N_0(W, \epsilon)$  and  $l \leq 2NW(1 - \epsilon)$ . Here  $C_2(W, \epsilon)$  and  $N_0(N, \epsilon)$  are constants specified in Lemma 2.3.

**Proof** (of Corollary I.1).

The main approach is to bound  $\tilde{\mathfrak{s}}_{N,W}^{(l)}(f)$ ,  $W \leq |f| \leq 1/2$  with the expressions presented in Lemma H.1. Suppose  $\epsilon \in (0, 1)$  is fixed.

1. For fixed  $l$  and large  $N$ :

In order to quantify the decay rate of  $|\tilde{\mathfrak{s}}_{N,W}^{(l)}(f)|$ , we exploit some results concerning of  $f_4(f)$  from [32] and  $f_5(f)$  as follows:

$$|J_0(x)| \leq 1, \quad \forall x \geq 0, \quad (32)$$

and for any  $\frac{\arccos(A-N^{-3/2})}{2\pi} \leq |f| \leq 1/2$ , one may verify that

$$\begin{aligned} |f_5(f)| &\leq \frac{1}{((A - \cos(2\pi f))(1 - \cos(2\pi f)))^{1/4}} \\ &\leq \frac{1}{((A - (A - N^{-3/2}))(1 - (A - N^{-3/2})))^{1/4}} \\ &\leq \frac{1}{((N^{-3/2})(N^{-3/2}))^{1/4}} = N^{3/4}, \end{aligned}$$

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  $c_5 \sim \sqrt{2\sqrt{2\alpha} (1 - \lambda_{N,W}^{(l)})}$ . Plugging these into Lemma H.1 and utilizing Lemma 2.3, we get the exponential decay of  $|\tilde{\mathfrak{s}}_{N,W}^{(l)}(f)|$ ,  $|f| \geq W$  as

$$|\tilde{\mathfrak{s}}_{N,W}^{(l)}(f)| \leq \begin{cases} C'_7(W, \epsilon) N^{1/2} e^{-\frac{C_2}{2} N}, & W \leq |f| \leq \arccos(A - N^{-3/2})/2\pi, \\ C'_8(W, \epsilon) N^{3/4} e^{-\frac{C_2}{2} N}, & \arccos(A - N^{-3/2})/2\pi \leq |f| \leq 1/2, \end{cases}$$

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 2.3.

2. For large  $N$  and  $l = \lfloor 2NW(1 - \epsilon') \rfloor$ ,  $\forall \epsilon' \in (0, \epsilon]$ :

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 (32),  $d_4 = \pi^{1/2} (1 - A^2)^{-1/4} 2^{-1/2} N^{1/2} d_6$  and  $d_6 \sim \sqrt{\frac{1 - \lambda_{N,W}^{(l)}}{2\pi}}$ .

Plugging these into Lemma H.1 and utilizing the bound on  $\lambda_{N,W}^{(l)}$  in Lemma 2.3, we get the exponential decay of  $|\tilde{\mathfrak{s}}_{N,W}^{(l)}(f)|$ ,  $|f| \geq W$  as

$$|\tilde{\mathfrak{s}}_{N,W}^{(l)}(f)| \leq \begin{cases} C''_7(W, \epsilon) N^{1/2} e^{-\frac{C_2}{2} N}, & W \leq |f| \leq \arccos[A - N^{-1}]/2\pi, \\ C''_8(W, \epsilon) N^{1/2} e^{-\frac{C_2}{2} N}, & \arccos[A - N^{-1}]/2\pi \leq |f| \leq 1/2, \end{cases}$$

for all  $l = \lfloor 2NW(1 - \epsilon') \rfloor$ ,  $\forall \epsilon' \in (0, \epsilon]$  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 2.3.

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 I.1.  $\square$

**Lemma I.2.** ([37]) For fixed  $W \in (0, \frac{1}{2})$  and  $\epsilon \in (0, \frac{1}{2W} - 1)$ ,  $\tilde{\mathbf{s}}_{N,W}^{(l)}(f)$  and  $\tilde{\mathbf{s}}_{N, \frac{1}{2}-W}^{(N-1-l)}(f)$  satisfy

$$|\tilde{\mathbf{s}}_{N,W}^{(l)}(f)| = |\tilde{\mathbf{s}}_{N, \frac{1}{2}-W}^{(N-1-l)}(\frac{1}{2} - f)|$$

for all  $l \geq 2NW(1 + \epsilon)$ .

Now we can conclude that  $\langle \mathbf{e}_f, \mathbf{s}_{N,W}^{(l)} \rangle$  decays exponentially in  $N$  for all  $l \geq 2NW(1 + \epsilon)$  and  $|f| \leq W$  by combining the above results.

**Corollary I.3.** 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

$$|\langle \mathbf{e}_f, \mathbf{s}_{N,W}^{(l)} \rangle| = |\tilde{\mathbf{s}}_{N,W}^{(l)}(f)| \leq C_6(W', \epsilon') N^{3/4} e^{-\frac{C_2(W', \epsilon')}{2} N}, \quad \forall |f| \leq W$$

for all  $N \geq N_0(W', \epsilon')$  and all  $l \geq 2NW(1 + \epsilon)$ . Here,  $C_2(W', \epsilon')$  and  $N_0(W', \epsilon')$  are constants specified in Lemma 2.3 with respect to  $W'$  and  $\epsilon'$ , and  $C_6(W', \epsilon')$  is the constant specified in Corollary I.1 with respect to  $W'$  and  $\epsilon'$ .

**Proof of Corollary I.3.** Let  $l' = N - 1 - l$ . For all  $l \geq 2NW(1 + \epsilon)$ , we have

$$l' = N - 1 - l \leq N - 2NW(1 + \epsilon) = 2N(\frac{1}{2} - W)(1 - \frac{W}{\frac{1}{2} - W}\epsilon).$$

Let  $W' = \frac{1}{2} - W$  and  $\epsilon' = \frac{W}{\frac{1}{2}-W}\epsilon \in (0, 1)$ . It follows from Corollary I.1 and Lemma I.2 that

$$|\langle \mathbf{e}_f, \mathbf{s}_{N,W}^{(l)} \rangle| = |\langle \mathbf{e}_{\frac{1}{2}-f}, \mathbf{s}_{N,W'}^{(l')} \rangle| \leq C_6(W', \epsilon') N^{3/4} e^{-\frac{C_2(W', \epsilon')}{2} N}, \quad \forall |f| \leq W$$

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

Recall that  $C_6(W', \epsilon') = \max\left\{\pi^{1/2} \left(\frac{2}{\alpha}\right)^{1/4}, 2^{-1}(1 - A^2)^{-1/4}\right\} \sqrt{C_1(W', \epsilon')}$  with  $A = \cos(2\pi W)$  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  $(1 - A^2)^{-1/4} \rightarrow 1/\sqrt{2\pi W}$  as  $W \rightarrow 0$ . Also we have  $\left(\frac{2}{\alpha}\right)^{1/4} \rightarrow 1/\sqrt{\pi W}$  as  $W \rightarrow 0$ . 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

$$\begin{aligned} \|\mathbf{e}_f - \mathbf{P}_{[S_{N,W}]_k} \mathbf{e}_f\|_2^2 &= \sum_{l=2NW(1+\epsilon)}^{N-1} |\langle \mathbf{e}_f, \mathbf{s}_{N,W}^{(l)} \rangle|^2 \\ &\leq \sum_{l=2NW(1+\epsilon)}^{N-1} C_6^2(W', \epsilon') N^{3/2} e^{-C_2(W', \epsilon') N} \\ &\leq C_9(W', \epsilon') N^{5/2} e^{-C_2(W', \epsilon') N} \end{aligned}$$

for all  $|f| \leq W$  and  $N \geq N_0(W', \epsilon')$ , where  $C_9(W', \epsilon') = C_6^2(W', \epsilon')$ .  $\square$

## J Proof of Corollary 3.10

**Proof.** Suppose  $f \in [f_i - W_i, f_i + W_i]$  for some particular  $i \in [J]$ . Let  $C_{10}(\mathbb{W}, \epsilon) = \max\{C_9(W'_i, \epsilon'), \forall i \in [J]\}$  and  $C_{11}(\mathbb{W}, \epsilon) = \min\{C_2(W'_i, \epsilon'), \forall i \in [J]\}$ . It follows from Theorem 3.9 that

$$\begin{aligned} \|\mathbf{e}_f - \mathbf{P}_{\Psi} \mathbf{e}_f\|_2^2 &\leq \|\mathbf{e}_f - \mathbf{P}_{[E_{f_i} S_{N,W_i}]_{2NW_i(1+\epsilon)}} \mathbf{e}_f\|_2^2 \\ &= \|\mathbf{e}_{f-f_i} - \mathbf{P}_{[S_{N,W_i}]_{2NW_i(1+\epsilon)}} \mathbf{e}_{f-f_i}\|_2^2 \\ &\leq C_9(W'_i, \epsilon') N^{5/2} e^{-C_2(W'_i, \epsilon') N} \leq C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N} \end{aligned}$$

for all  $N \geq N_0(W'_i, \epsilon')$ . We complete the proof by setting  $N_2(\mathbb{W}, \epsilon) = \max\{N_0(W'_i, \epsilon'), \forall i \in [J]\}$ .  $\square$

## K Proof of Theorem 3.11

**Proof.** Since  $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{J-1}$  are independent and zero-mean, we have

$$\mathbb{E} [\|\mathbf{x}\|_2^2] = \sum_{n=0}^{N-1} \mathbb{E} [|\mathbf{x}[n]|^2] = \sum_{n=0}^{N-1} \sum_{0 \leq i, i' \leq J-1} \mathbb{E} [\mathbf{x}_i[n] \overline{\mathbf{x}_{i'}[n]}] = \sum_{n=0}^{N-1} \sum_{i=0}^{J-1} \mathbb{E} [|\mathbf{x}_i[n]|^2] = N \sum_{i=0}^{J-1} \frac{1}{J} = N.$$

Applying Theorem 2.4, we acquire

$$\mathbb{E} \left[ \left\| \mathbf{x}_i - \mathbf{P}_{[\mathbf{E}_{f_i} \mathbf{S}_{N, W_i}]_{k_i}} \mathbf{x}_i \right\|_2^2 \right] = \frac{1}{|\mathbb{W}|} \sum_{l=k_i}^{N-1} \lambda_{N, W_i}^{(l)}.$$

Note that the power spectrum  $P_{x_i}(F)$  assumed in (20) results in the constant  $\frac{1}{|\mathbb{W}|}$  instead of  $\frac{1}{2W_i}$ . Now, we have

$$\begin{aligned} \mathbb{E} [\|\mathbf{x} - \mathbf{P}_{\Psi} \mathbf{x}\|_2^2] &= \mathbb{E} \left[ \left\| \sum_{i=0}^{J-1} \mathbf{x}_i - \mathbf{P}_{\Psi} \left( \sum_{i=0}^{J-1} \mathbf{x}_i \right) \right\|_2^2 \right] = \mathbb{E} \left[ \left\| \sum_{i=0}^{J-1} (\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i) \right\|_2^2 \right] \\ &= \mathbb{E} \left[ \left( \sum_{i=0}^{J-1} (\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i)^H \right) \left( \sum_{i=0}^{J-1} (\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i) \right) \right] \\ &= \mathbb{E} \left[ \sum_{i=0}^{J-1} \|\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i\|_2^2 + \sum_{i=0}^{J-1} \sum_{i'=0, i' \neq i}^{J-1} (\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i)^H (\mathbf{x}_{i'} - \mathbf{P}_{\Psi} \mathbf{x}_{i'}) \right] \\ &= \sum_{i=0}^{J-1} \mathbb{E} [\|\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i\|_2^2] + \sum_{i=0}^{J-1} \sum_{i'=0, i' \neq i}^{J-1} \mathbb{E} [(\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i)^H (\mathbf{x}_{i'} - \mathbf{P}_{\Psi} \mathbf{x}_{i'})] \\ &= \sum_{i=0}^{J-1} \mathbb{E} [\|\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i\|_2^2] + \sum_{i=0}^{J-1} \sum_{i'=0, i' \neq i}^J \mathbb{E} [\mathbf{x}_i^H \mathbf{x}_{i'} - \mathbf{x}_i^H \mathbf{P}_{\Psi} \mathbf{x}_{i'}] \\ &= \sum_{i=0}^{J-1} \mathbb{E} [\|\mathbf{x}_i - \mathbf{P}_{\Psi} \mathbf{x}_i\|_2^2] \leq \sum_{i=0}^{J-1} \mathbb{E} \left[ \left\| \mathbf{x}_i - \mathbf{P}_{[\mathbf{E}_{f_i} \mathbf{S}_{N, W_i}]_{k_i}} \mathbf{x}_i \right\|_2^2 \right] \\ &= \sum_{i=0}^{J-1} \frac{1}{|\mathbb{W}|} \sum_{l=k_i}^{N-1} \lambda_{N, W_i}^{(l)} \end{aligned}$$

where the equality in the sixth line follows because  $\mathbb{E} [\mathbf{x}_i^H \mathbf{x}_i] = (\mathbb{E} [\mathbf{x}_{i'}])^H (\mathbb{E} [\mathbf{x}_i]) = 0$  and  $\mathbb{E} [\mathbf{x}_{i'}^H \mathbf{P}_{\Psi} \mathbf{x}_i] = (\mathbb{E} [\mathbf{x}_{i'}])^H (\mathbb{E} [\mathbf{P}_{\Psi} \mathbf{x}_i]) = 0$  for all  $i', i \in [J], i' \neq i$ , and the inequality in the sixth line follows because the column space of  $[\mathbf{E}_{f_i} \mathbf{S}_{N, W_i}]_{k_i}$  is inside the column space of  $\Psi$  for all  $i \in [J]$ .  $\square$

## L Proof of Corollary 3.12

**Proof.** It is useful to express the sampled bandpass signal  $\mathbf{x}$  as

$$\mathbf{x} = \int_{\mathbb{W}} \tilde{x}(f) \mathbf{e}_f df, \quad (33)$$

where we recall that  $\tilde{x}(f)$  denotes the DFT 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 (33) that

$$\begin{aligned} \|\mathbf{x} - \mathbf{P}_{\Psi} \mathbf{x}\|_2^2 &= \left\| \int_{\mathbb{W}} \tilde{x}(f) \mathbf{e}_f df - \int_{\mathbb{W}} \tilde{x}(f) \mathbf{P}_{\Psi} \mathbf{e}_f df \right\|_2^2 \\ &= \left\| \int_{\mathbb{W}} \tilde{x}(f) (\mathbf{e}_f - \mathbf{P}_{\Psi} \mathbf{e}_f) df \right\|_2^2 \\ &\leq \int_{\mathbb{W}} |\tilde{x}(f)|^2 df \cdot \int_{\mathbb{W}} \|\mathbf{e}_f - \mathbf{P}_{\Psi} \mathbf{e}_f\|_2^2 df \\ &\leq \int_{\mathbb{W}} |\tilde{x}(f)|^2 df \cdot C_{10}(\mathbb{W}, \epsilon) N^{5/2} e^{-C_{11}(\mathbb{W}, \epsilon) N}, \end{aligned}$$

where the third line follows from the Cauchy-Schwarz inequality and the last line follows from (19) and the fact that  $\int_{\mathbb{W}} \|\mathbf{e}_f - \mathbf{P}_{\Psi} \mathbf{e}_f\|_2^2 df \leq |\mathbb{W}| \sup_{f \in \mathbb{W}} \|\mathbf{e}_f - \mathbf{P}_{\Psi} \mathbf{e}_f\|_2^2 \leq \sup_{f \in \mathbb{W}} \|\mathbf{e}_f - \mathbf{P}_{\Psi} \mathbf{e}_f\|_2^2$ .  $\square$

## References

- [1] F. Ahmad, Q. Jiang, and M. G. 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.
- [2] R. G. Baraniuk and P. Steeghs. Compressive radar imaging. In *Proc. 2007 IEEE Radar Conference*, pages 128–133, April 2007.
- [3] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, 2008.
- [4] 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.
- [5] 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.
- [6] E. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.
- [7] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. Pure Appl. Math.*, 67(6):906–956, 2014.
- [8] 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.
- [9] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] M. A. Davenport and M. B. Wakin. Compressive sensing of analog signals using Discrete Prolate Spheroidal Sequences. *Appl. Comput. Harmon. Anal.*, 33(3):438–472, 2012.
- [14] G. Davis. *Adaptive nonlinear approximations*. PhD thesis, Courant Institute of Mathematical Sciences New York, 1994.
- [15] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.
- [16] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*, 41(3):613–627, 1995.
- [17] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.
- [18] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization. *Proc. Natl. Acad. Sci.*, 100(5):2197–2202, 2003.
- [19] A. Eftekhari, J. Romberg, and M. B. Wakin. Matched filtering from limited frequency samples. *IEEE Trans. Inf. Theory*, 59(6):3475–3496, 2013.
- [20] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Imaging Sci.*, 5(1):179–202, 2012.
- [21] F. A. Grünbaum. Toeplitz matrices commuting with tridiagonal matrices. *Linear Algebra Appl.*, 40(0):25 – 36, 1981.
- [22] J. A. Hogan and J. D. Lakey. *Duration and Bandwidth Limiting: Prolate Functions, Sampling, and Applications*. Springer Science & Business Media, 2011.
- [23] 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.

- [24] R. A. Horn and C. R. Johnson, editors. *Matrix Analysis*. Cambridge University Press, New York, NY, USA, 1986.
- [25] S. Izu and J. D. Lakey. Time-frequency localization and sampling of multiband signals. *Acta Appl. Math.*, 107(1-3):399–435, 2009.
- [26] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [27] 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.
- [28] H. Landau. On the density of phase-space expansions. *IEEE Trans. Inf. Theory*, 39(4):1152–1156, 1993.
- [29] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 3rd edition, 2008.
- [30] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.
- [31] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmonic Anal.*, 26(3):301–321, 2009.
- [32] 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.
- [33] A. Papoulis. A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Trans. Circuits Syst.*, 22(9):735–742, 1975.
- [34] L. Reichel and L. N. Trefethen. Eigenvalues and pseudo-eigenvalues of Toeplitz matrices. *Linear Algebra Appl.*, 162:153–185, 1992.
- [35] E. Sejdić, A. Can, L. F. Chaparro, C. M. 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 Process.*, 2012(1):1–14, 2012.
- [36] 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)*, pages 2849–2852, March 2008.
- [37] D. Slepian. Prolate Spheroidal Wave Functions, Fourier analysis, and uncertainty. V- The discrete case. *Bell Syst. Tech. J.*, 57(5):1371–1430, 1978.
- [38] H. Stark and J. W. Woods. *Probability, Random Processes, and Estimation Theory for Engineers*. Prentice-Hall, 1986.
- [39] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Trans. Inf. Theory*, 59(11):7465–7490, 2013.
- [40] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.
- [41] 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.
- [42] T. Zemen and C. F. Mecklenbräuker. Time-variant channel estimation using Discrete Prolate Spheroidal Sequences. *IEEE Trans. Signal Process.*, 53(9):3597–3607, 2005.
- [43] T. Zemen, C. F. Mecklenbräuker, F. Kaltenberger, and B. H. Fleury. Minimum-energy band-limited predictor with dynamic subspace selection for time-variant flat-fading channels. *IEEE Trans. Signal Process.*, 55(9):4534–4548, 2007.
- [44] T. Zemen and A. F. Molisch. Adaptive reduced-rank estimation of nonstationary time-variant channels using subspace selection. *IEEE Trans. Veh. Technol.*, 61(9):4042–4056, 2012.
- [45] Z. Zhu and M. B. Wakin. Wall clutter mitigation and target detection using Discrete Prolate Spheroidal Sequences. In *3rd Int. Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, June 2015.