TIME-LIMITED TOEPLITZ OPERATORS ON ABELIAN GROUPS: APPLICATIONS IN INFORMATION THEORY AND SUBSPACE APPROXIMATION

ZHIHUI ZHU AND MICHAEL B. WAKIN

Abstract. Toeplitz operators are fundamental and ubiquitous in signal processing and information theory as models for linear, time-invariant (LTI) systems. Due to the fact that any practical system can access only signals of finite duration, time-limited restrictions of Toeplitz operators are naturally of interest. To provide a unifying treatment of such systems working on different signal domains, we consider time-limited Toeplitz operators on locally compact abelian groups with the aid of the Fourier transform on these groups. In particular, we survey existing results concerning the relationship between the spectrum of a time-limited Toeplitz operator and the spectrum of the corresponding non-time-limited Toeplitz operator. We also develop new results specifically concerning the eigenvalues of time-frequency limiting operators on locally compact abelian groups. Applications of our unifying treatment are discussed in relation to channel capacity and in relation to representation and approximation of signals.

1. Introduction

This paper deals with generalizations of certain concepts from elementary signals and systems analysis, which we first review.

1.1. Spectral analysis of linear, time-invariant systems.

Linear, time-invariant (LTI) systems are ubiquitous in signal processing and control theory, and it is well known that the output of a continuous-time (CT) LTI system with input signal $x(t)$ can be computed using the convolution integral

$$y(t) = (x * h)(t) = \int_{\tau=-\infty}^{\infty} h(t-\tau)x(\tau)\,\mathrm{d}\,\tau, \tag{1}$$

where $h(t) \in L_2(\mathbb{R})$ is the impulse response of the system. Such a system can equivalently be viewed as a linear operator $\mathcal{H} : L_2(\mathbb{R}) \to L_\infty(\mathbb{R})$, where

$$\mathcal{H}(x)(t) = \int_{\tau=-\infty}^{\infty} h(t-\tau)x(\tau)\,\mathrm{d}\,\tau. \tag{2}$$

Because this linear operator involves a kernel function $h(t-\tau)$ that depends only on the difference $t-\tau$, we refer to it as a *Toeplitz operator.*[1] In this setting, the behavior of the Toeplitz operator can be naturally understood in the frequency domain: for an input signal $x(t)$ with continuous-time Fourier transform (CTFT)

$$\widehat{x}(F) = \int_{t=-\infty}^{\infty} x(t)e^{-j2\pi Ft}\,\mathrm{d}\,t, \forall\ F \in \mathbb{R}, \tag{3}$$

2010 *Mathematics Subject Classification.* 47B35, 47N70, 94A12.

Key words and phrases. Toeptliz operators, time-frequency analysis, subspace approximation, eigenvalue distribution.

[1]Our notion of Toeplitz operators follows from the definition of Toeplitz operators in [24, Section 7.2].

the CTFT of the output signal $y(t)$ will satisfy $\widehat{y}(F) = \widehat{x}(F)\widehat{h}(F)$, where $\widehat{h}(F)$ denotes the CTFT of the impulse response $h(t)$ and is also known as the *frequency response* of the system. The spectrum of the Toeplitz operator $\mathcal{H}$ also coincides with $\widehat{h}(F)$ [24, Section 7.2]. Note that the spectrum of a linear operator $\mathcal{H}$, a generalization of the set of eigenvalues of a matrix, is the set of complex numbers λ such that $\mathcal{H} - \lambda\mathcal{I}$ (where $\mathcal{I}$ denotes the identity operator) is not invertible.

Similar facts hold for discrete-time (DT) LTI systems, where the response to an input signal $x[n]$ is given by the convolution

$$y[n] = \sum_{m=-\infty}^{\infty} h[n-m]x[m], \tag{4}$$

where $h[n] \in \ell_2(\mathbb{Z})$ is the impulse response of the DT system. Such a system can equivalently be viewed as a linear operator $H : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$, which corresponds to multiplication of the input signal $x \in \ell_2(\mathbb{Z})$ by the bi-infinite Toeplitz matrix

$$H = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & h[0] & h[-1] & h[-2] & \ddots \\ \ddots & h[1] & h[0] & h[-1] & \ddots \\ \ddots & h[2] & h[1] & h[0] & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{5}$$

We note that $H[m,n] = h[m-n]$ for all $m, n \in \mathbb{Z}$. The behavior of this system can also be interpreted as multiplication in the discrete-time Fourier transform (DTFT) domain where the DTFT of the impulse response $h[n]$ is defined as:

$$\widehat{h}(f) = \sum_{n=-\infty}^{\infty} h[n]e^{-j2\pi fn}. \tag{6}$$

The spectrum of H also coincides with $\widehat{h}(f)$ [24].

1.2. **The effects of time-limiting.** Practical systems do not have access to input or output signals of infinite duration, which motivates the study of time-limited versions of LTI systems. Consider for example the situation where a CT system zeros out an input signal outside the interval $[0,T]$. (Or similarly, the system may pad with zeros an input signal that was originally compactly supported on $[0,T]$.) The system then computes the convolution shown in (2) and finally time-limits the output signal to the same interval $[0,T]$. For such a situation we may define a new linear operator $\mathcal{H}_T : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ (a "time-limited" version of $\mathcal{H}$), where

$$\mathcal{H}_T(x)(t) = \begin{cases} \int_{\tau=0}^{T} h(t-\tau)x(\tau)\,\mathrm{d}\,\tau, & t \in [0,T] \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

An analogous time-limited version of H (from (5)) may be defined for DT systems. Supposing that the input and output signals are time-limited to the index set $\{0, 1, \ldots, N-1\}$, we define the $N \times N$ Toeplitz matrix[2] $\boldsymbol{H}_N$ as

$$\boldsymbol{H}_N[m,n] = h[m-n], \quad \forall\ 0 \le m, n \le N-1. \tag{8}$$

Such a matrix can also be viewed as a linear operator on $\mathbb{C}^N$.

[2]Through the paper, finite-dimensional vectors and matrices are indicated by bold characters and we index such vectors and matrices beginning at 0.

A natural question is: *What effect do the time-limiting operations have on the system behavior?* More precisely, how similar is the spectrum of $\mathcal{H}_T$ to that of $\mathcal{H}$, and in what sense do the eigenvalues of $\mathcal{H}_T$ converge to the frequency response $\widehat{h}(F)$ as $T \to \infty$? Here $\mathcal{H}_T$ is compact and thus has a discrete spectrum containing what we refer to as its eigenvalues; the number of eigenvalues is countable by the spectral theorem for compact operators [9]. Analogously, how similar is the spectrum of $\boldsymbol{H}_N$ to that of H, and in what sense do the eigenvalues of $\boldsymbol{H}_N$ converge to the frequency response $\widehat{h}(f)$ as $N \to \infty$? As we discuss, the answers to such questions provide insight into matters such as the capacity (or effective bandwidth) of time-limited communication channels and the number of degrees of freedom (or effective dimensionality) of certain related signal families. Answering these questions relies on deeper insight into the spectrum of Toeplitz operators.

1.2.1. *Toeplitz and time-limited Toeplitz operators.* In this paper, we distinguish between Toeplitz operators (such as $\mathcal{H}$ and H) and time-limited Toeplitz operators such as $\mathcal{H}_T$ and $\boldsymbol{H}_N$.[3] We focus primarily on Toeplitz and time-limited Toeplitz operators that are Hermitian, i.e., $h(-t) = h^*(t)$ for $\mathcal{H}$ and $\mathcal{H}_T$ and $h[-n] = h^*[n]$ for H and $\boldsymbol{H}_N$.

We note that finite size Toeplitz *matrices* (such as $\boldsymbol{H}_N$) are of considerable interest in statistical signal processing and information theory [22, 24, 33, 49, 53]. The covariance matrix of a random vector obtained by sampling a wide-sense stationary (WSS) random process is an example of such a matrix. More general Toeplitz operators have been extensively studied since O. Toeplitz and C. Carathéodory [6, 75]; see [24] for a very comprehensive review. Time-limited convolutions are also important in machine learning and computer vision. As an example, modern convolutional neural networks (CNNs)—which have demonstrated excellent performance in numerous computer vision tasks [40]—contain large numbers of convolutional layers, each of which mainly involves two-dimensional convolution (that can be written as a doubly block circulant matrix, which is approximately Toeplitz) applied to the input.

Unfortunately, there are no simple formulas for the eigenvalues of time-limited Toeplitz operators such as $\mathcal{H}_T$ and $\boldsymbol{H}_N$. This stands in contrast to the operators $\mathcal{H}$ and H, whose spectrum was given simply by the frequency response of the corresponding LTI system. Notably, although the discrete Fourier transform (DFT) is the canonical tool for frequency analysis in $\mathbb{C}^N$, the DFT basis vectors (complex exponentials of the form $e^{j2\pi\frac{nk}{N}}$ with $k \in \{0, 1, \ldots, N-1\}$) do not, in general, constitute eigenvectors of the matrix $\boldsymbol{H}_N$, unless this matrix is *circulant* in addition to being Toeplitz. Consequently, the spectrum of $\boldsymbol{H}_N$ cannot in general be obtained by taking the DFT of the time-limited impulse response $\{h[0], h[1], \ldots, h[N-1]\}$.

Fortunately, in many applications it is possible to relate the eigenvalues of a time-limited Toeplitz operator to the eigenvalues of the original (non-time-limited) Toeplitz operator, thus guaranteeing that certain essential behavior of the original system is preserved in its time-limited version. We discuss these connections, as well as their applications, in more detail in the following subsections.

1.2.2. *Time-frequency limiting operators.* Shannon introduced the fundamental concept of capacity in the context of communication in [64], in which we find the answers to questions such as the capacity of a CT band-limited channel which operates substantially limited to a time interval $[0, T]$. In [64], the answer was derived in a probabilistic setting, while another notation of ϵ-capacity was introduced by Kolmogorov in [74] for approaching a similar question in the deterministic setting of signal (or functional) approximation. The functional approximation approach was further investigated by Landau, Pollak, and Slepian, who wrote a series of seminal papers exploring the degree to which a band-limited signal can be *approximately* time-limited [43, 44, 65, 67, 69]. Recently, Lim

[3] Our usage of these terms is consistent with the terminology in [24, Section 7.2], although in that work time-limited Toeplitz operators are referred to as *finite* Toeplitz operators.

and Franceschcetti [46, 47] provided a connection between Shannon's capacity from the probabilistic setting and Kolmogorov's capacity from the deterministic setting when communication occurs using band-limited functions.

To give a precise description, consider the case of a CT Toeplitz operator $\mathcal{H}$ (as in (2)) that corresponds to an *ideal low-pass filter*. That is, $\mathcal{H} = \mathcal{B}_W$, where $\mathcal{B}_W : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is a band-limiting operator that takes the CTFT of an input function on $L_2(\mathbb{R})$, sets it to zero outside $[-W, W]$ and then computes the inverse CTFT. The impulse response of this system is given by the sinc function $h(t) = \frac{\sin(2\pi W t)}{\pi W t}$, and the frequency response of this system $\widehat{h}(F)$ is simply the indicator function of the interval $[-W, W]$.

Similarly, define $\mathcal{T}_T : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ to be a time-limiting operator that sets a function to zero outside $[0, T]$, and finally consider the time-limited Toeplitz operator $\mathcal{H}_T = \mathcal{T}_T \mathcal{H} \mathcal{T}_T = \mathcal{T}_T \mathcal{B}_W \mathcal{T}_T$. Observe that $\mathcal{H}_T$ can be viewed as a composition of time- and band-limiting operators.

The eigenvalues of $\mathcal{T}_T \mathcal{B}_W \mathcal{T}_T$ were extensively investigated in [43, 69], which discuss the "lucky accident" that $\mathcal{T}_T \mathcal{B}_W \mathcal{T}_T$ commutes with a certain second-order differential operator whose eigenfunctions are a special class of functions—the *prolate spheroidal wave functions* (PSWFs).

The eigenvalues of the corresponding composition of time- and band-limiting operators in the discrete case, a Toeplitz matrix $\boldsymbol{H}_N$ whose entries are samples of a digital sinc function, were studied by Slepian in [67]. The eigenvectors of this matrix are time-limited versions of the *discrete prolate spheroidal sequences* (DPSSs) which, as we discuss further in Section 4.4, provide a highly efficient basis for representing sampled band-limited signals and have proved to be useful in numerous signal processing applications.

In both the CT and DT settings, the eigenvalues of the time-limited Toeplitz operator exhibit a particular behavior: when sorted by magnitude, there is a cluster of eigenvalues close to (but not exceeding) 1, followed by an abrupt transition, after which the remaining eigenvalues are close to 0. This crudely resembles the rectangular shape of the frequency response of the original band-limiting operator. Moreover, the number of eigenvalues near 1 is approximately equal to the time-frequency area (which equals $2TW$ in the CT example above). More details on these facts, including a complete treatment of the DT case, are provided in Section 4.

1.2.3. *Szegő's theorem.* For more general Toeplitz operators—beyond ideal low-pass filters—the eigenvalues of the corresponding time-limited Toeplitz operators can be described using Szegő's theorem.

We describe this in the DT case. Consider a DT Hermitian Toeplitz operator H which corresponds to a system with frequency response $\widehat{h}(f)$, as described in (5) and (6). For $N \in \mathbb{N}$, let $\boldsymbol{H}_N$ denote the resulting time-limited Hermitian Toeplitz operator, as in (8), and let the eigenvalues of $\boldsymbol{H}_N$ be arranged as $\lambda_0(\boldsymbol{H}_N) \geq \cdots \geq \lambda_{N-1}(\boldsymbol{H}_N)$. Suppose $\widehat{h} \in L_\infty([0,1])$. Then Szegő's theorem [24] describes the collective asymptotic behavior (as $N \to \infty$) of the eigenvalues of the sequence of Hermitian Toeplitz matrices $\{\boldsymbol{H}_N\}$ as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=0}^{N-1} \vartheta(\lambda_l(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widehat{h}(f)) df, \tag{9}$$

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

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

In words, this says that as $N \to \infty$, the average eigenvalue of $\boldsymbol{H}_N$ converges to the average value of the frequency response $\widehat{h}(f)$ of the original Toeplitz operator H. As a second example, suppose

$\widehat{h}(f) > 0$ (and thus $\lambda_l(\boldsymbol{H}_N) > 0$ for all $l \in \{0, 1, \ldots, N-1\}$ and $N \in \mathbb{N}$) and let ϑ be the log function. Then Szegő's theorem indicates that

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

This relates the determinants of the Toeplitz matrices $\boldsymbol{H}_N$ to the frequency response $\widehat{h}(f)$ of the original Toeplitz operator H.

Szegő's theorem has been widely used in the areas of signal processing, communications, and information theory. A paper and review by Gray [22,23] serve as a remarkable elementary introduction in the engineering literature and offer a simplified proof of Szegő's original theorem. The result has also been extended in several ways. For example, the Avram-Parter theorem [3, 52] relates the collective asymptotic behavior of the singular values of a general (non-Hermitian) Toeplitz matrix to the magnitude response $|\widehat{h}(f)|$. Tyrtyshnikov [76] proved that Szegő's theorem holds if $\widehat{h}(f) \in \mathbb{R}$ and $\widehat{h}(f) \in L^2([0,1])$, and Zamarashkin and Tyrtyshnikov [83] further extended Szegő's theorem to the case where $\widehat{h}(f) \in \mathbb{R}$ and $\widehat{h}(f) \in L^1([0,1])$. Sakrison [56] extended Szegő's theorem to higher dimensions. Gazzah et al. [20] and Gutiérrez-Gutiérrez and Crespo [26] extended Gray's results on Toeplitz and circulant matrices to block Toeplitz and block circulant matrices and derived Szegő's theorem for block Toeplitz matrices.

Similar results also hold in the CT case, with the operators $\mathcal{H}$ and $\mathcal{H}_T$ as defined in (2) and (7). Let $\lambda_\ell(\mathcal{H}_\mathrm{T})$ denote the ℓ^{th}-largest eigenvalue of $\mathcal{H}_T$. Suppose $\widehat{h}(F)$ is a real-valued, bounded and integrable function, i.e., $\widehat{h}(F) \in \mathbb{R}$, $\widehat{h}(F) \in L^\infty(\mathbb{R})$, and $\widehat{h}(F) \in L^1(\mathbb{R})$. Then Szegő's theorem in the continuous case [24] states that the eigenvalues of $\mathcal{H}_T$ satisfy

$$\lim_{T\to\infty} \frac{\#\{\ell : a < \lambda_\ell(\mathcal{H}_T) < b\}}{T} = \left|\{F : a < \widehat{h}(F) < b\}\right| \tag{10}$$

for any interval (a, b) such that $|\{F : \widehat{h}(F) = a\}| = |\{F : \widehat{h}(F) = b\}| = 0$. Here $|\cdot|$ denotes the length (or Lebesgue measure) of an interval. Stated differently, this result implies that the eigenvalues of the operator $\mathcal{H}_T$ have asymptotically the same distribution as the values of $\widehat{h}(F)$ when F is distributed with uniform density along the real axis.

We remark that although the collective behavior of the eigenvalues of the time-frequency limiting operators discussed in Section 1.2.2 can be characterized using Szegő's theorem, finer results on the eigenvalues have been established for this special class of time-limited Toeplitz operators. We discuss Szegő's theorem for general operators in Section 3, and we discuss results for time-frequency limiting operators in Section 4.

1.2.4. *Time-limited Toeplitz operators on locally compact abelian groups.* One of the important pieces of progress in harmonic analysis made in last century is the definition of the Fourier transform on locally compact abelian groups [55]. This framework for harmonic analysis on groups not only unifies the CTFT, n-dimensional CTFT, DTFT, and DFT (for signal domains, or groups, corresponding to $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{Z}$, and $\mathbb{Z}_N := \{0, 1, \ldots, N-1\}$, respectively), but it also allows these transforms to be generalized to other signal domains. This, in turn, makes possible the analysis of new applications such as steerable principal component analysis (PCA) [78] where the domain is the rotation angle on $[0, 2\pi)$, an imaging system with a pupil of finite size [15], line-of-sight (LOS) communication systems with orbital angular momentum (OAM)-based orthogonal multiplexing techniques [80], and many other applications such as those involving rotations in three dimensions [8, Chapter 5].

In this paper, we consider the connections between Toeplitz and time-limited[4] Toeplitz operators on locally compact abelian groups. As we review in Section 2, one important fact carries over from the classical setting described in Section 1.1: the eigenvalues of any Toeplitz operator on a locally compact abelian group are given by the generalized frequency response of the system.

In light of this fact, we are once again interested in questions such as: How does the spectrum of a time-limited Toeplitz operator relate to the spectrum of the original (non-time-limited) Toeplitz operator? In what sense do the eigenvalues converge as the domain of time-limiting approaches the entire group? The answers to such questions will provide new insight into the effective dimensionality of certain signal families (such as the class of signals that are time-limited and essentially band-limited) and the amount of information that can be transmitted in space or time by band-limited functions.

1.3. **Contribution and paper organization.** This paper focuses on the spectra of time-limited Toeplitz operators and the resulting implications in signal processing and information theory, containing part survey and part novel work. In particular, as new results, we study the spectra of time-frequency limiting operators on locally compact abelian groups and analyze applications in representation and approximation of band-limited signals, generalizing the existing results in Section 1.2.2.

The remainder of paper is organized as follows. Section 2 reviews harmonic analysis on locally compact abelian groups and draws a connection between time-limited Toeplitz operators and the effective dimensionality of certain related signal families. Next, Section 3 reviews Szegő's theorem and its (existing) generalization to locally compact abelian groups. Applications are discussed in channel capacity, signal representation, and numerical analysis. Finally, Section 4 reviews existing results on the eigenvalues of time-frequency limiting operators and generalizes these results to locally compact abelian groups. New applications of this unifying treatment are discussed in relation to channel capacity and in relation to representation and approximation of signals. This work also opens up new questions concerning applications and research directions, which we discuss at the ends of Section 3 and Section 4.

2. Preliminaries

We briefly introduce some notation used throughout the paper. Sets (of variables, functions, etc.) are denoted in blackboard font as $\mathbb{A}, \mathbb{B}, \ldots$. Operators are denoted in calligraphic font as $\mathcal{A}, \mathcal{B}, \ldots$.

2.1. Harmonic analysis on locally compact abelian groups.

2.1.1. *Groups and dual groups.* Let $\mathbb{G}$ (with a binary operation $\circ$) denote a *locally compact abelian group*, which can be either discrete or continuous, and either compact or non-compact. A *character* $\chi_\xi : \mathbb{G} \to \mathbb{T}$ of $\mathbb{G}$ is a continuous group homomorphism from $\mathbb{G}$ with values in the circle group $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ satisfying

$$|\chi_\xi(g)| = 1,\ \chi_\xi^*(g) = \chi_\xi(g^{-1}),\ \chi_\xi(h \circ g) = \chi_\xi(h)\chi_\xi(g),$$

for any $g, h \in \mathbb{G}$. Here $\chi_\xi^*(g)$ is the complex conjugate of $\chi_\xi(g)$. The set of all characters on $\mathbb{G}$ introduces a locally compact abelian group, called the *dual group* of $\mathbb{G}$ and denoted by $\widehat{\mathbb{G}}$ if we pair $(g, \xi) \to \chi_\xi(g)$ for all $\xi \in \widehat{\mathbb{G}}$ and $g \in \mathbb{G}$. In most references the character is denoted simply by χ rather than by χ_ξ. However, we use here the notation χ_ξ in order to emphasize that the character can be viewed as a function of two elements $g \in \mathbb{G}$ and $\xi \in \widehat{\mathbb{G}}$, and for any $\xi \in \widehat{\mathbb{G}}$, χ_ξ is a function of g. In this sense, $\chi_\xi(g)$ can be regarded as the value of the character χ_ξ evaluated at the

[4]Here we use "time" to be consistent with the preceding discussion, but this concept broadly applies to other domains such as the spatial domain.

group element g. Table 1 lists several examples of groups $\mathbb{G}$, along with the corresponding binary operation $\circ$ and dual group $\widehat{\mathbb{G}}$, that have relevance in signal processing and information theory. Here $\mathrm{mod}(a,b) = \frac{a}{b} - \lfloor \frac{a}{b} \rfloor$, where $\lfloor c \rfloor$ is the largest integer that is not greater than c.

TABLE 1. Examples of groups $\mathbb{G}$, along with their dual groups $\widehat{\mathbb{G}}$ and Fourier transforms (FT). Below, CT denotes continuous time, DT denotes discrete time, FS denotes Fourier series, and DFT denotes the discrete Fourier transform.

group $\mathbb{G}$	dual group $\widehat{\mathbb{G}}$	g	binary operation $g_1 \circ g_2$	ξ	$\chi_\xi(g)$	FT
$\mathbb{R}$	$\mathbb{R}$	t	$t_1 + t_2$	F	$e^{j2\pi Ft}$	CTFT
$\mathbb{R}^n$	$\mathbb{R}^n$	$\boldsymbol{t}$	$\boldsymbol{t}_1 + \boldsymbol{t}_2$	$\boldsymbol{F}$	$e^{j2\pi\langle \boldsymbol{F},\boldsymbol{t}\rangle}$	CTFT in $\mathbb{R}^n$
unit circle $[0,1)$	$\mathbb{Z}$	t	$\mathrm{mod}(t_1 + t_2, 1)$	k	$e^{j2\pi tk}$	CTFS
$\mathbb{Z}$	unit circle	n	$n_1 + n_2$	f	$e^{j2\pi fn}$	DTFT
$\mathbb{Z}_N = N$ roots of unity	$\mathbb{Z}_N = N$ roots of unity	n	$\mathrm{mod}(n_1 + n_2, N)$	k	$e^{j2\pi \frac{nk}{N}}$	DFT

2.1.2. *Fourier transforms.* The characters $\{\chi_\xi\}_{\xi\in\widehat{\mathbb{G}}}$ play an important role in defining the Fourier transform for functions in $L_2(\mathbb{G})$. In particular, the Pontryagin duality theorem [55], named after Lev Semennovich Pontryagin who laid the foundation for the theory of locally compact abelian groups, generalizes the conventional CTFT on $L_2(\mathbb{R})$ and CT Fourier series for periodic functions to functions defined on locally compact abelian groups.

Theorem 1 (Pontryagin duality theorem [55])**.** *Let $\mathbb{G}$ be a locally compact abelian group and μ be a Haar measure on $\mathbb{G}$. Let $x(g) \in L_1(\mathbb{G})$. Then the Fourier transform $\widehat{x}(\xi)$ is defined by $\widehat{x}(\xi) = \int_{\mathbb{G}} x(g)\chi_\xi^*(g)\,\mathrm{d}\,\mu(g)$. For each Haar measure μ on $\mathbb{G}$ there is a unique Haar measure ν on $\widehat{\mathbb{G}}$ such that the following inverse Fourier transform holds $x(g) = \int_{\widehat{\mathbb{G}}} \widehat{x}(\xi)\chi_\xi(g)\,\mathrm{d}\,\nu(\xi)$. The Fourier transform satisfies Parseval's theorem: $\int_{\mathbb{G}} |x(g)|^2\,\mathrm{d}\,\mu(g) = \int_{\widehat{\mathbb{G}}} |\widehat{x}(\xi)|^2\,\mathrm{d}\,\nu(\xi)$.*

With certain technical tricks, the same Fourier transform and Parseval's theorem in Theorem 1 can be extended to $L_2(\mathbb{G})$; see [72, Theorem 59], [55, 1.6.1] and the references therein for details. Only Haar measures and integrals are considered throughout this paper. We note that the unique Haar measure ν on $\widehat{\mathbb{G}}$ depends on the choice of Haar measure μ (which is unique except for positive scaling factors) on $\mathbb{G}$. We illustrate this point with the conventional DFT as an example where $g = n \in \mathbb{G} = \mathbb{Z}_N$, $\xi = k \in \widehat{\mathbb{G}} = \mathbb{Z}_N$, and $\chi_\xi(g) = e^{j2\pi\frac{nk}{N}}$. If we choose the counting measure (where each element of $\mathbb{G}$ receives a value of 1) on $\mathbb{G}$, then we must use the normalized counting measure (where each element of $\widehat{\mathbb{G}}$ receives a value of $\frac{1}{N}$) on $\widehat{\mathbb{G}}$. The DFT and inverse DFT become

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

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

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

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

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

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

For each group $\mathbb{G}$ and dual group $\widehat{\mathbb{G}}$ listed in Table 1, the table also includes the corresponding Fourier transform.

2.1.3. *Convolutions.* For any $x(g), y(g) \in L_2(\mathbb{G})$, we define the convolution between them by

$$(x \star y)(g) := \int_{\mathbb{G}} y(h)x(h^{-1} \circ g) \, \mathrm{d}\, h. \tag{11}$$

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

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

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

Similar to the fact that Toeplitz operators (2) and Toeplitz matrices (5) are closely related to the convolutions in Section 1.1, the convolution (11) can be viewed as a linear operator $\mathcal{X} : L_2(\mathbb{G}) \to L_\infty(\mathbb{G})$ that computes the convolution between the input function $y(g)$ and $x(g)$:

$$(\mathcal{X}y)(g) = \int_{\mathbb{G}} x(h^{-1} \circ g) y(h) \, \mathrm{d}\, h.$$

We refer to $\mathcal{X}$ as a *Toeplitz operator* since this linear operator involves a kernel function $x(h^{-1} \circ g)$ that depends only on the difference $h^{-1} \circ g$. We call $\widehat{x}(\xi)$, the Fourier transform of $x(g)$, the *symbol* corresponding to the Toeplitz operator $\mathcal{X}$.

Finally, let $\mathbb{A} \in \mathbb{G}$ be a subset of $\mathbb{G}$. As explained in Section 1.2, we are also interested in the *time-limited Toeplitz operator*[5] $\mathcal{X}_\mathbb{A} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$, where

$$(\mathcal{X}_\mathbb{A} y)(g) = \begin{cases} \int_\mathbb{A} x(h^{-1} \circ g) y(h) \, \mathrm{d}\, h, & g \in \mathbb{A}, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Recall that the spectrum of a linear operator $\mathcal{X}$ is the set of complex numbers λ such that $\mathcal{X} - \lambda\mathcal{I}$ (where $\mathcal{I}$ denotes the identity operator) is not invertible. Here the time-limited Toeplitz operator $\mathcal{X}_\mathbb{A}$ is compact and thus has a discrete spectrum containing what refer to as its eigenvalues. There is no simple formula for exactly expressing the eigenvalues of $\mathcal{X}_\mathbb{A}$. Instead, we are interested in questions such as: How does the spectrum of the time-limited Toeplitz operator $\mathcal{X}_\mathbb{A}$ relate to the spectrum of the original (non-time-limited) Toeplitz operator $\mathcal{X}$? In what sense do the eigenvalues converge as the domain $\mathbb{A}$ of time-limiting approaches the entire group $\mathbb{G}$? We discuss answers to these questions in Section 3 and Section 4.

2.2. **The effective dimensionality of a signal family.** One of the useful applications of characterizing the spectrum of time-limited Toeplitz operators is in computing the effective dimensionality (or the number of degrees of freedom) of certain related signal families. In this section, we formalize this notion of effective dimensionality for a set of functions defined on a group $\mathbb{G}$.

[5] $\mathcal{X}_\mathbb{A}$ is also referred to as a Toeplitz operator in [24, 27, 39, 48]. Again, we note that here "time" refers to the domain in $\mathbb{G}$.

2.2.1. *Definitions.* We begin by defining general sets of time-limited functions that we are interested in; in later sections we discuss these functions in the context of communications and signal processing. Specifically, suppose $\mathbb{A}$ is a subset of $\mathbb{G}$ and let $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)) \subset L_2(\mathbb{A})$ denote the set of functions controlled by a symbol $\widehat{\phi}(\xi)$:

$$\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)) := \left\{ x \in L_2(\mathbb{A}) : x(g) = \int_{\widehat{\mathbb{G}}} \alpha(\xi)\widehat{\phi}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi, \int |\alpha(\xi)|^2\,\mathrm{d}\,\xi \leq 1, g \in \mathbb{A} \right\}, \tag{13}$$

which is a subset of $L_2(\mathbb{A})$. We note that in (13), the symbol $\widehat{\phi}(\xi)$ is fixed and we discuss its role soon.

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

(14)

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

where $\mathrm{P}_{\mathbb{M}_n} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ represents the orthogonal projection onto the subspace $\mathbb{M}_n$. We define the width $d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n)$ of the set $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ with respect to the subspace $\mathbb{M}_n$ as follows:

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

which also represents the largest distance from the elements in $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ to the n-dimensional subspace $\mathbb{M}_n$. The Kolmogorov n-width [38], $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$ of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ in $L_2(\mathbb{G})$ is defined as the smallest width $d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n)$ over all n-dimensional subspaces of $L_2(\mathbb{G})$; that is

$$d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) := \inf_{\mathbb{M}_n} d(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \mathbb{M}_n). \tag{15}$$

In summary, the n-width $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$ characterizes how well the set $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ can be approximated by an n-dimensional subspace of $L_2(\mathbb{G})$. By its definition, $d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)))$ is non-increasing in terms of the dimensionality n. For any fixed $\epsilon > 0$, we define the *effective dimensionality*, or *number of degrees of freedom*, of the set $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ at level ϵ as [19]

$$\mathcal{N}(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \epsilon) = \min \left\{ n : d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) < \epsilon \right\}. \tag{16}$$

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

We note that the reason we impose an energy constraint on the elements x of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ in (13) is that we use the absolute distance to quantify the proximity of x to the subspace $\mathbb{M}_n$ in (14).

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

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

The adjoint $\mathcal{A}^* : L_2(\mathbb{A}) \to L_2(\widehat{\mathbb{G}})$ is given by $(\mathcal{A}^* x)(\xi) = \int_{\mathbb{A}} x(g)\widehat{\phi}^*(\xi)\chi_\xi^*(g)\,\mathrm{d}\,g$. The composition of $\mathcal{A}$ and $\mathcal{A}^*$ gives a self-adjoint operator $\mathcal{A}\mathcal{A}^* : L_2(\mathbb{A}) \to L_2(\mathbb{A})$ as follows:

$$
\begin{aligned}
(\mathcal{A}\mathcal{A}^* x)(g) &= \int_{\widehat{\mathbb{G}}} \widehat{\phi}(\xi)\chi_\xi(g) \int_{\mathbb{A}} x(h)\widehat{\phi}^*(\xi)\chi_\xi^*(h)\,\mathrm{d}\,h\,\mathrm{d}\,\xi \\
&= \int_{\mathbb{A}} x(h) \int_{\widehat{\mathbb{G}}} \left|\widehat{\phi}(\xi)\right|^2 \chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi\,\mathrm{d}\,h = \int_{\mathbb{A}} x(h)(\phi \star \phi^*)(h^{-1}\circ g)\,\mathrm{d}\,h,
\end{aligned} \tag{17}
$$

where $\phi(g) = \int_{\widehat{\mathbb{G}}} \widehat{\phi}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi$ is the inverse Fourier transform of $\widehat{\phi}$. In words, compared with (12), the self-adjoint operator $\mathcal{A}\mathcal{A}^*$ can be viewed as a time-limited Toeplitz operator with the kernel $\phi \star \phi^*$.

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

Proposition 1. *[54] Let the eigenvalues of $\mathcal{A}\mathcal{A}^*$ be denoted and arranged as $\lambda_1 \geq \lambda_2 \geq \cdots$. Then the n-width of $\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))$ can be computed as*

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

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

The proof of Proposition 1 is given in Appendix A.

3. General Toeplitz Operators on Locally Abelian Groups

Let $x(g) \in L_2(\mathbb{G})$, whose Fourier transform is given by $\widehat{x}(\xi)$. Now we are well equipped to consider the eigenvalue distribution of a general time-limited Toeplitz operator $\mathcal{X}_\mathbb{A}$ (which is formally defined in (12)) on a locally abelian group; in particular we are interested in the relationship between the spectrum of of the time-limited Toeplitz operator $\mathcal{X}_\mathbb{A}$ and $\widehat{x}(\xi)$. The operator $\mathcal{X}_\mathbb{A}$ is completely continuous and its eigenvalues are denoted by $\lambda_\ell(\mathcal{X}_\mathbb{A})$. Before presenting the main results, we introduce new notation for subsets of $\mathbb{G}$ (or $\widehat{\mathbb{G}}$) which are asymptotically increasing to cover the whole group. This is similar to how we discussed the cases where $T \to \infty$ and $N \to \infty$ in Section 1.2. To that end, let $\mathbb{A}_\tau, \tau \in (0, \infty)$ be a system of Borel subsets of $\mathbb{G}$ with boundaries of measure zero such that $0 < \mu(\mathbb{A}_\tau) < \infty$. The subscript τ is sometimes dropped when it is clear from the context. We can view $\mathbb{A}_\tau$ as a set of subsets that depend on the parameter τ. One can also define a system of subsets with multiple parameters.

3.1. Generalized Szegő's theorem.

Abundant effort [24, 27, 39, 48] has been devoted to extending the conventional Szegő's theorem for a general time-limited Toeplitz operator $\mathcal{X}_\mathbb{A}$. Let $\mathcal{N}(\mathcal{X}_\mathbb{A}; (a, b)) =: \#\{\ell : a < \lambda_\ell(\mathcal{X}_\mathbb{A}) < b\}$ denote the number of eigenvalues of $\mathcal{X}_\mathbb{A}$ that are between a and b. We summarize the following generalized Szegő's theorem concerning the collective behavior of the eigenvalues of $\mathcal{X}_\mathbb{A}$ and relating them to $\widehat{x}(\xi)$ (the spectrum of the corresponding non-time-limited operator $\mathcal{X}$).

Theorem 2 (Generalized Szegő's theorem [24, 27, 39, 48])**.** *Let $x(g) \in L_2(\mathbb{G})$ and $\mathcal{X}_{\mathbb{A}_\tau}$ be the time-limited Toeplitz operator defined in (12). Suppose the following holds almost everywhere:*

$$
\lim_{\tau\to\infty} \mathbb{A}_\tau = \mathbb{G}.
$$

Then for all intervals (a, b) such that $\nu(\{\xi : \widehat{x}(\xi) = a\}) = \nu(\{\xi : \widehat{x}(\xi) = b\}) = 0$, we have

$$
\lim_{\tau\to\infty} \frac{\mathcal{N}(\mathcal{X}_{\mathbb{A}_\tau}; (a, b))}{\mu(\mathbb{A}_\tau)} = \nu(\{\xi : a < \widehat{x}(\xi) < b\}). \tag{18}
$$

In a nutshell, Theorem 2 implies that the eigenvalues of the time-limited Toeplitz operator $\mathcal{X}_{\mathbb{A}_\tau}$ are closely related to $\widehat{x}(\xi)$, the spectrum of the corresponding non-time-limited Toeplitz operator $\mathcal{X}$. Some work instead presents (18) as

$$\lim_{\tau\to\infty} \frac{\mathcal{N}(\mathcal{X}_{\mathbb{A}_\tau}; [a,b])}{\mu(\mathbb{A}_\tau)} = \nu\left(\{\xi : a \le \widehat{x}(\xi) \le b\}\right). \tag{19}$$

One can understand the equivalence between (18) and (19) as the boundary of the interval makes no difference since $\nu\left(\{\xi : \widehat{x}(\xi) = a\}\right) = 0$ and $\nu\left(\{\xi : \widehat{x}(\xi) = b\}\right) = 0$.

In words, Theorem 2 implies that the eigenvalue distribution of the operator $\mathcal{X}_{\mathbb{A}_\tau}$ asymptotically converges to the distribution of the Fourier transform of $x(g)$. We now compare Theorem 2 with the conventional Szegő's theorems in Section 1.2.3 that have widely appeared in information theory and signal processing. We note that (18) has exactly the same form as (10) for the time-limited operator $\mathcal{H}_T$ in (7).

For the Toeplitz matrix $\boldsymbol{H}_N$ defined in (8), at first glance, (9) is slightly different than what is expressed in (18) which implies

$$\lim_{N\to\infty} \frac{\mathcal{N}(\boldsymbol{H}_N; (a,b))}{N} = \left|\left\{f : f \in [0,1), a < \widehat{h}(f) < b\right\}\right| \tag{20}$$

for all intervals (a,b) such that $\left|\left\{f : \widehat{h}(f) = a\right\}\right| = 0$ and $\left|\left\{f : \widehat{h}(f) = b\right\}\right| = 0$. In fact, (9) and (20) are equivalent if we view $\widehat{h} : [0,1) \to \mathbb{R}$ as a random variable and define $\boldsymbol{\lambda}_{\boldsymbol{H}_N} : \{0,1,\dots,N-1\} \to \mathbb{R}$ to be a discrete random variable by $\boldsymbol{\lambda}_{\boldsymbol{H}_N}[\ell] = \lambda_\ell(\boldsymbol{H}_N)$. In probabilistic language, set

$$F_{\widehat{h}}(a) := \left|\left\{f : f \in [0,1), \widehat{h}(f) \le a\right\}\right|$$

to be the cumulative distribution function (CDF) associated to $\widehat{h}$. Also denote the CDF associated to $\boldsymbol{\lambda}_{\boldsymbol{H}_N}$ by

$$F_{\boldsymbol{\lambda}_{\boldsymbol{H}_N}}(a) := \frac{\mathcal{N}\left(\boldsymbol{H}_N; (-\infty, a]\right)}{N} = \frac{\#\left\{\ell : \lambda_\ell(\boldsymbol{H}_N) \le a\right\}}{N}.$$

The following result, known as the Portmanteau lemma, gives two equivalent descriptions of weak convergence in terms of the CDF and the means of the random variables.

Lemma 1. *[77, Portmanteau lemma] The following are equivalent:*

(1) $\lim_{N\to\infty} \frac{1}{N}\sum_{\ell=0}^{N-1} \vartheta(\lambda_\ell(\boldsymbol{H}_N)) = \int_0^1 \vartheta(\widehat{h}(f))df$ *for all bounded, continuous functions* ϑ*;*
(2) $\lim_{N\to\infty} F_{\boldsymbol{\lambda}_{\boldsymbol{H}_N}}(a) = F_{\widehat{h}}(a)$ *for every point* a *at which* $F_{\widehat{h}}$ *is continuous.*

Note that if $F_{\widehat{h}}$ is continuous at a, then $\left|\left\{f : \widehat{h}(f) = a\right\}\right| = 0$. Thus the equivalence between (9) and (20) follows from the Portmanteau lemma. In words, (20) implies that certain collective behaviors of the eigenvalues of each Toeplitz matrix are reflected by the symbol $\widehat{h}(f)$.

We note that (9) is one of the descriptions of weak convergence of a sequence of random variables in the Portmanteau lemma [77] (also see Lemma 1). Thus, throughout the paper, we also refer to the collective behavior (like that characterized by (9)) of the eigenvalues as the *distribution* of the eigenvalues.

In the following subsections, we discuss applications of the generalized Szegő's theorem.

3.2. **Application: Subspace approximations.**

3.2.1. *Convolutions with a pulse.* We first consider the set of functions obtained by time-limiting the convolution between $\alpha(g)$ and a fixed function $\phi(g)$:

$$\mathbb{W}(\mathbb{A},\phi(g)) := \left\{ x \in L_2(\mathbb{A}) : x(g) = \int_{\mathbb{G}} \alpha(h)\phi(h^{-1}\circ g)\,\mathrm{d}\,h, \int_{\mathbb{G}} |\alpha(g)|^2\,\mathrm{d}\,g \le 1, g \in \mathbb{A} \right\}. \tag{21}$$

We note that $\mathbb{W}(\mathbb{A},\phi(g))$ is equivalent to $\mathbb{W}(\mathbb{A},\widehat{\phi}(\xi))$ defined in (13) by rewriting $x(g)$ in (21):

$$x(g) = \int_{\mathbb{G}} \alpha(h)\phi(h^{-1}\circ g)\,\mathrm{d}\,h = \int_{\widehat{\mathbb{G}}} \widehat{\alpha}(\xi)\widehat{\phi}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi,$$

which is exactly the same form of $x(g)$ in (13). This model (21) arises in radar signal processing, channel sensing, and super-resolution of pulses through an unknown channel. Proposition 1 implies that the n-width of $\mathbb{W}(\mathbb{A},\phi(g))$ is given by

$$d_n(\mathbb{W}(\mathbb{A},\phi(g))) = \sqrt{\lambda_n(\mathcal{A}\mathcal{A}^*)},$$

where $\mathcal{A}\mathcal{A}^*$ defined in (17) is time-limited Toeplitz operator with the kernel $\phi\star\phi^*$. Now Theorem 2 along with (16) reveals the effective dimensionality, or number of degrees of freedom, of the set $\mathbb{W}(\mathbb{A},\phi(g))$ at level ϵ as follows.

Corollary 1. *Suppose* $\lim_{\tau\to\infty}\mathbb{A}_\tau = \mathbb{G}$ *holds almost everywhere. Then for any* $\epsilon > 0$ *such that* $\nu\left(\left\{\xi : \widehat{\phi}(\xi) = \epsilon\right\}\right) = 0$, *we have*

$$\lim_{\tau\to\infty} \frac{\mathcal{N}(\mathbb{W}(\mathbb{A}_\tau,\phi(g)),\epsilon)}{\mu(\mathbb{A}_\tau)} = \nu\left(\left\{\xi : \left|\widehat{\phi}(\xi)\right| > \epsilon\right\}\right).$$

Proof. By (16), we have $\mathcal{N}(\mathbb{W}(\mathbb{A}_\tau,\phi(g)),\epsilon) = \min\left\{n : \sqrt{\lambda_n} < \epsilon\right\} = \#\left\{n : \sqrt{\lambda_n} \ge \epsilon\right\}$. Note that the Fourier transform of $\phi\star\phi^*$ is $\left|\widehat{\phi}(\xi)\right|^2$. Then Corollary 1 then follows directly by applying Theorem 2 to $\mathcal{A}\mathcal{A}^*$ (which is a time-limited Toeplitz operator with the kernel $\phi\star\phi^*$). □

3.2.2. *Shifts of a signal.* We now consider a slightly different model where the function of interest is a linear combination of continuous shifts of a given signal $\phi(g) \in L_2(\mathbb{G})$:

$$\mathbb{S}(\mathbb{A},\phi(g)) := \left\{ x \in L_2(\mathbb{A}) : x(g) = \int_{\mathbb{A}} \alpha(h)\phi(h^{-1}\circ g)\,\mathrm{d}\,h, \int_{\mathbb{A}} |\alpha(g)|^2\,\mathrm{d}\,g \le 1 \right\}. \tag{22}$$

Define $\mathcal{T}_{\mathbb{A}} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ as a time-limiting operator that makes a function zero outside $\mathbb{A}$. We can also rewrite $x(g) = \int_{\mathbb{G}} (\mathcal{T}_{\mathbb{A}}\alpha)(h)\phi(h^{-1}\circ g)\,\mathrm{d}\,h = (\mathcal{T}_{\mathbb{A}}\alpha)\star\phi$ for $g \in \mathbb{A}$, the convolution between the time-limited function $(\mathcal{T}_{\mathbb{A}}\alpha)(g)$ and $\phi(g)$. By zero-padding the signal x outside $\mathbb{A}$, we may also rewrite it simply as $x = \mathcal{T}_{\mathbb{A}}((\mathcal{T}_{\mathbb{A}}\alpha)\star\phi)$. Analogously, we could express the function $x(g)$ in (21) as $x = \mathcal{T}_{\mathbb{A}}(\alpha\star\phi)$. Now it is clear the model for $\mathbb{W}(\mathbb{A},\phi(g))$ and the one for $\mathbb{S}(\mathbb{A},\phi(g))$ differs in the location of the time-limiting operator $\mathcal{T}_{\mathbb{A}}$.

To investigate the effective dimension or the number of degrees of freedom for $\mathbb{S}(\mathbb{A},\phi(g))$, we define the operator $\mathcal{S} : L_2(\mathbb{A}) \to L_2(\mathbb{G})$ as:

$$(\mathcal{S}\alpha)(g) = \int_{\mathbb{A}} \alpha(h)\phi(h^{-1}\circ g)\,\mathrm{d}\,h,\ g \in \mathbb{G}.$$

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

$$(\mathcal{S}^* x)(g) = \int_{\mathbb{G}} \phi^*(g^{-1}\circ h)x(h)\,\mathrm{d}\,h,\ g \in \mathbb{A}.$$

We then have the self-adjoint operator $\mathcal{S}^*\mathcal{S} : L_2(\mathbb{A}) \to L_2(\mathbb{A})$

$$(\mathcal{S}^*\mathcal{S}\alpha)(g) = \int_{\mathbb{G}} \phi^*(g^{-1} \circ h) \int_{\mathbb{A}} \alpha(\eta)\phi(\eta^{-1} \circ h) \,\mathrm{d}\,\eta \,\mathrm{d}\,h$$
$$= \int_{\mathbb{A}} \int_{\mathbb{G}} \phi^*(g^{-1} \circ h)\phi(\eta^{-1} \circ h) \,\mathrm{d}\,h\alpha(\eta) \,\mathrm{d}\,\eta = \int_{\mathbb{A}} r(\eta^{-1} \circ g)\alpha(\eta) \,\mathrm{d}\,\eta,$$

where $r(g) := \int_{\mathbb{G}} \phi(h)\phi(g^{-1} \circ h) \,\mathrm{d}\,h$ is the autocorrelation function of the function ϕ. Thus, $\mathcal{S}^*\mathcal{S}$ is a Toeplitz operator of the form (12). Similar to Proposition 1, we can study the effective dimension of the set of shifted signals in (22) by looking at the eigenvalue distribution of the self-adjoint operator $\mathcal{S}\mathcal{S}^*$. Note that in this case $\mathcal{S}\mathcal{S}^*$ is not a Toeplitz operator, but it has the same nonzero eigenvalues as $\mathcal{S}^*\mathcal{S}$. Thus, we can exploit the eigenvalue distribution of $\mathcal{S}^*\mathcal{S}$ to infer the number of degrees of freedom for the set $\mathbb{S}(\mathbb{A}, \phi(g))$. This is formally established in the following result.

Proposition 2. *Let the eigenvalues of $\mathcal{S}^*\mathcal{S}$ be denoted and arranged as $\lambda_1 \geq \lambda_2 \geq \cdots$. Then the n-width of $\mathbb{S}(\mathbb{A}, \phi(g))$ can be computed as $d_n(\mathbb{S}(\mathbb{A}, \phi(g))) = \sqrt{\lambda_n}$, and the optimal n-dimensional subspace to represent $\mathbb{S}(\mathbb{A}, \phi(g))$ is the subspace spanned by the first n eigenvectors of $\mathcal{S}\mathcal{S}^*$.*

Finally, Theorem 2 implies that the eigenvalue distribution of the Toeplitz operator $\mathcal{S}^*\mathcal{S}$ is asymptotically equivalent to $\widehat{r}(\xi) = \int_{\mathbb{G}} r(g)\chi_\xi(g) \,\mathrm{d}\,g$, the power spectrum of ϕ if we view r as the autocorrelation of ϕ.

3.3. **Application: Eigenvalue estimation.** In many applications such as spectrum sensing algorithm for cognitive radio [86], it is desirable to understand the *individual* asymptotic behavior of the eigenvalues of time-limited Toeplitz operators rather than the collective behavior of the eigenvalues provided by Szegő's theorem (Theorem 2). As a special case, efficiently estimating the spectral norm (i.e., the largest singular value) of Toeplitz matrices is crucial in certain applications. For example, the Lipschitz constant of a CNN has wide implications in understanding the key properties of the neural network such as its generalization and robustness. Unfortunately, computing the exact Lipschitz constant of a neural network is known to be NP-hard [58]. Recent work [2, 82] proposed methods for computing upper bounds of the Lipschitz constant for each layer (and hence for the entire network) by efficiently estimating the spectral norm of the corresponding (block) Toeplitz matrices. We will review related recent progress on characterizing the individual behavior of the eigenvalues for Toeplitz matrices. To our knowledge, the individual behavior of the eigenvalues has only recently been investigated for Toeplitz matrices.

Bogoya et al. [4] studied the individual asymptotic behavior of the eigenvalues of Toeplitz matrices by interpreting Szegő's theorem in (9) in probabilistic language and related the eigenvalues to the values obtained by sampling $\widehat{h}(f)$ uniformly in frequency on $[0, 1)$:

$$\lim_{N\to\infty} \max_{0\leq\ell\leq N-1} \left| \lambda_\ell(\boldsymbol{H}_N) - \widehat{h}(\frac{\rho(\ell)}{N}) \right| = 0 \tag{23}$$

if the range of $\widehat{h}(f)$ is connected. Here $\widehat{h}(\frac{\rho(\ell)}{N})$ is the permuted form of $\widehat{h}(\frac{\ell}{N})$ such that $\widehat{h}(\frac{\rho(0)}{N}) \geq \widehat{h}(\frac{\rho(1)}{N}) \geq \cdots \geq \widehat{h}(\frac{\rho(\ell)}{N})$. Thus, if the symbol $\widehat{h}(f)$ is known, we can sample it uniformly to get reasonable estimates for the eigenvalues of the Toeplitz matrix.

Despite the power of Szegő's theorem, in many scenarios (such as certain coding and filtering applications [22, 53]), one may only have access to $\boldsymbol{H}_N$ and not $\widehat{h}$. In such cases, it is still desirable to have practical and efficiently computable estimates of the individual eigenvalues of $\boldsymbol{H}_N$. We recently showed [93] that we can construct a certain sequence of $N \times N$ circulant matrices such that the eigenvalues of the circulant matrices asymptotically converge to those of the Toeplitz matrices. Transforming the Toeplitz matrix into a circulant matrix can be performed extremely efficiently

using closed form expressions; the eigenvalues of the circulant matrix can then be computed very efficiently (in $O(N \log N)$ using the fast Fourier transform (FFT)).

When the sequence $h[n]$ is not symmetric about the origin, the Avram-Parter theorem [3, 52], a generalization of Szegő's theorem, relates the collective asymptotic behavior of the singular values of a general (non-Hermitian) Toeplitz matrix to the absolute value of its symbol, i.e., $|\widehat{h}(f)|$. Bogoya et al. [4] also showed that the singular values of $\boldsymbol{H}_N$ asymptotically converge to the uniform samples of $|\widehat{h}(f)|$ provided the range of the symbol $|\widehat{h}(f)|$ is connected.

3.4. **Questions.** Inspired by the applications listed above, we raise two questions concerning the generalized Szegő's theorem (Theorem 2). The first question concerns the individual behavior of the eigenvalues.

Question 1. *Is it possible to extend the result* (23) *concerning the individual behavior of the eigenvalues for Toeplitz matrices to general Toeplitz operators?*

We note that both the conventional Szegő's theorem listed in Section 1.2.3 and the generalized Szegő's theorem (Theorem 2) characterize asymptotic behavior of the eigenvalues.

Question 2. *Is it possible to establish a non-asymptotic result concerning the eigenvalue behavior (either collective or individual) for the general Toeplitz operators* $\mathcal{X}_{\mathbb{A}_\tau}$*?*

4. Time-Frequency Limiting Operators on Locally Compact Abelian Groups

In this section, we consider a special case of time-limited Toeplitz operators: time-frequency limiting operators on locally compact abelian groups, to be formally defined soon. As we have briefly explained in Section 1.2, time-frequency limiting operators in the context of the classical groups where $\mathbb{G}$ are the real-line, $\mathbb{Z}$, and $\mathbb{Z}_N$ play important roles in signal processing and communication. By considering time-frequency limiting operators on locally compact abelian groups, we aim to (i) provide a unified treatment of the previous results on the eigenvalues of the operators resulting in PSWFs, DPSSs, and periodic DPSSs (PDPSSs) [25, 32]; and (ii) extend these results to other signal domains such as rotations in a plane and three dimensions [8, Chapter 5]. In particular, we will investigate the eigenvalues of time-frequency limiting operators on locally compact abelian groups and show that they exhibit similar behavior to both the conventional CT and DT settings: when sorted by magnitude, there is a cluster of eigenvalues close to (but not exceeding) 1, followed by a relatively sharp transition, after which the remaining eigenvalues are close to 0. This behavior also resembles the rectangular shape of the frequency response of the original band-limiting operator. Although this collective behavior can be characterized using Szegő's theorem, finer results (particularly non-asymptotic results) on the eigenvalues have been established in special cases, which we will review in detail. We will also discuss the applications of this unifying treatment in relation to channel capacity and to representation and approximation of signals.

To introduce the time-frequency limiting operators, consider two subsets $\mathbb{A} \in \mathbb{G}$ and $\mathbb{B} \in \widehat{\mathbb{G}}$. Recall that $\mathcal{T}_{\mathbb{A}} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ is a time-limiting operator that makes a function zero outside $\mathbb{A}$. Also define $\mathcal{B}_{\mathbb{B}} = \mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F} : L_2(\mathbb{G}) \to L_2(\mathbb{G})$ as a band-limiting operator that takes the Fourier transform of an input function on $L_2(\mathbb{G})$, sets it to zero outside $\mathbb{B}$, and then computes the inverse Fourier transform. The operator $\mathcal{B}_{\mathbb{B}}$ acts on $L_2(\mathbb{G})$ as a convolutional integral operator:

$$(\mathcal{B}_{\mathbb{B}}x)(g) = \int_{\mathbb{B}} \widehat{x}(\xi)\chi_\xi(g)\,\mathrm{d}\,\xi = \int_{\mathbb{B}} \left(\int_{\mathbb{G}} x(h)\chi_\xi^*(h)\,\mathrm{d}\,h \right) \chi_\xi(g)\,\mathrm{d}\,\xi = \int_{\mathbb{G}} K_{\mathbb{B}}(h^{-1} \circ g)x(h)\,\mathrm{d}\,h,$$

where

$$K_{\mathbb{B}}(h^{-1} \circ g) = \int_{\mathbb{B}} \chi_\xi^*(h)\chi_\xi(g)\,\mathrm{d}\,\xi = \int_{\mathbb{B}} \chi_\xi(h^{-1} \circ g)\,\mathrm{d}\,\xi. \tag{24}$$

It is of interest to study the eigenvalues of the following operators which we refer to as *time-frequency limiting operators*

$$\mathcal{O}_{\mathbb{A},\mathbb{B}} = \mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}, \text{ and } \quad \mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}. \tag{25}$$

Utilizing the expression for $\mathcal{B}_{\mathbb{B}}$, the operator $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ acts on any $x \in L_2(\mathbb{G})$ as follows

$$\left(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}x\right)(g) = \begin{cases} \int_{\mathbb{A}} K_{\mathbb{B}}(h^{-1}\circ g)x(h)\,\mathrm{d}\,h, & g \in \mathbb{A} \\ 0, & \text{otherwise.} \end{cases}$$

The operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ is symmetric and completely continuous and we denote its eigenvalues by $\lambda_\ell(\mathcal{O}_{\mathbb{A},\mathbb{B}})$. Due to the time- and band-limiting characteristics of the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$, the eigenvalues of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ are between 0 and 1. To see this, let $x(g) \in L_2(\mathbb{A})$:

$$\begin{aligned}\langle(\mathcal{O}_{\mathbb{A},\mathbb{B}}x)(g), x(g)\rangle &= \left\langle \int_{\mathbb{A}}\int_{\mathbb{B}} \chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi x(h)\,\mathrm{d}\,h, x(g)\right\rangle \\ &= \int_{\mathbb{B}}\left(\int_{\mathbb{A}}\int_{\mathbb{A}} \chi_\xi(h^{-1}\circ g)x(h)x^*(g)\,\mathrm{d}\,h\,\mathrm{d}\,g\right)\mathrm{d}\,\xi = \int_{\mathbb{B}} |\widehat{x}(\xi)|^2\,\mathrm{d}\,\xi \geq 0.\end{aligned}$$

On the other hand, we have $\int_{\mathbb{B}} |\widehat{x}(\xi)|^2\,\mathrm{d}\,\xi \leq \int_{\widehat{\mathbb{G}}} |\widehat{x}(\xi)|^2\,\mathrm{d}\,\xi = \int_{\mathbb{G}} |x(g)|^2\,\mathrm{d}\,g$.

4.1. **Eigenvalue distribution of time-frequency limiting operators.** To investigate the eigenvalues of the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}} = \mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$, we first note that without the time-limiting operator $\mathcal{T}_{\mathbb{A}}$, the eigenvalues of $\mathcal{B}_{\mathbb{B}}$ are simply given by the Fourier transform of $K_{\mathbb{B}}(g)$, and thus they are either 1 or 0. Our main question is how the spectrum of the time-frequency limiting operator relates to the spectrum of the band-limited operator. Based on the binary spectrum of $\mathcal{B}_{\mathbb{B}}$ and the intuition from Szegő's theorem in Theorem 2, we expect that the eigenvalues of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ to have a particular behavior: when sorted by magnitude, there should be a cluster of eigenvalues close to (but not exceeding) 1, followed by an abrupt transition, after which the remaining eigenvalues should be close to 0. Moreover, the number of effective (i.e., relatively large) eigenvalues should be essentially equal to the time-frequency area $|\mathbb{A}||\mathbb{B}|$. These results are confirmed below and reveal the dimensionality (or the number of degrees of freedom) of classes of band-limited signals observed over a finite time, which is fundamental to characterizing the performance limits of communication systems.

We note that similar to how we discussed the cases where $T \to \infty$ and $N \to \infty$ in Section 1.2, we will use $\mathbb{A}_\tau, \tau \in (0,\infty)$ to define the subsets of $\mathbb{G}$ that depend on τ. The subscript τ is often dropped when it is clear from the context. We now present one of our main results concerning the asymptotic behavior for the eigenvalues of the time-frequency limiting operators $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$ when $\mathbb{A}_\tau$ approaches $\mathbb{G}$.

Theorem 3. *Suppose $\mathbb{B}$ is a fixed subset of $\widehat{\mathbb{G}}$ and let $\epsilon \in (0, \frac{1}{2})$. Let*

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; (a,b)) := \#\left\{\ell : a < \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) < b\right\}$$

denote the number of eigenvalues of $\mathcal{T}_{\mathbb{A}_\tau}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}_\tau}$ that are between a and b. Then if

$$\lim_{\tau\to\infty} \mathbb{A}_\tau = \mathbb{G} \tag{26}$$

holds almost everywhere, we have

$$\sum_\ell \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{A}_\tau||\mathbb{B}|, \quad \sum_\ell \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{A}_\tau||\mathbb{B}| - o(|\mathbb{A}_\tau||\mathbb{B}|), \tag{27}$$

and

$$\lim_{\tau\to\infty} \frac{\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; [1-\epsilon, 1])}{|\mathbb{A}_\tau|} = |\mathbb{B}|, \quad \mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}; (\epsilon, 1-\epsilon)) = o\left(\frac{|\mathbb{A}_\tau||\mathbb{B}|}{\epsilon(1-\epsilon)}\right). \tag{28}$$

Here $|\cdot|$ denotes the Haar measure.

The proof of Theorem 3 is given in Appendix B. The limit in (26) is in the sense of convergence in measure on each compact set of $\mathbb{G}$. There are no specific shape constraints in $\mathbb{A}_\tau$ except that their boundaries are measure zero, though we note that in many cases of interest $\mathbb{A}_\tau$ is a closed set as in bandlimited signals [41], a union of closed sets as in multiband signals [92], or a scaling of a fixed set as in [19] where $\boldsymbol{Q}_\tau\mathbb{A} = \{\boldsymbol{Q}_\tau g : g \in \mathbb{A}\}$ where $\mathbb{G}$ is $\mathbb{R}^n$, $\mathbb{A}$ is a fixed set of $\mathbb{R}^n$ with boundary of measure zero, and $\boldsymbol{Q}_\tau \in \mathbb{R}^{n\times n}$ depends on τ such that $\lim_{\tau\to\infty} \boldsymbol{Q}_\tau\mathbb{A} = \mathbb{R}^n$. Theorem 3 formally confirms that the spectra of the time-frequency limiting operators resemble the rectangular shape of the spectrum of the band-limiting operator. As guaranteed by (28), the number of effective eigenvalues of the time-frequency limiting operator is asymptotically equal to the time-frequency area $|\mathbb{A}_\tau||\mathbb{B}|$. Similar results for time-frequency limiting operators in the context of classical groups where $\mathbb{G}$ is $\mathbb{R}^n$ are given in [19, 41]. We discuss the applications of Theorem 3 in channel capacity and representation and approximation of signals in more detail in the following subsections.

As mentioned before, the time-frequency limiting operators in the context of the classical groups where $\mathbb{G}$ are the real-line, $\mathbb{Z}$, and $\mathbb{Z}_N$ were first studied by Landau, Pollak, and Slepian who wrote a series of papers regarding the dimensionality of time-limited signals that are approximately band-limited (or vice versa) [43, 44, 65, 67, 69] (see also [66, 68] for concise overviews of this body of work). After that, a set of results concerning the number of eigenvalues within the transition region $(0, 1)$ have been established in [17, 30, 37, 45, 51, 87]. which will be reviewed in detail in the following remarks.

Remark 1. Using the explicit expressions for the character function $\chi_\xi(g)$ and the kernel $K_{\mathbb{B}}(g)$ and applying integration by parts for (39), one can improve the second term in (27) to $O(\log(|\mathbb{A}_\tau||\mathbb{B}|))$ for many common one-dimensional cases:

- Suppose $\mathbb{G} = \mathbb{R}$ and $\widehat{\mathbb{G}} = \mathbb{R}$. Let $\mathbb{A}_T = [-\frac{T}{2}, \frac{T}{2}]$ (where $\tau = T$ in (26)) and $\mathbb{B} = [-\frac{1}{2}, \frac{1}{2}]$ without loss of generality. Then the kernel $K_{\mathbb{B}}(t)$ turns out to be $K_{\mathbb{B}}(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi Ft}\,\mathrm{d}\,F = \frac{2\sin(\pi t)}{\pi t}$. Plugging in this form into (39) gives [28] $\sum_\ell(\lambda_\ell(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}))^2 = T - O(\log(T))$. In this case, the operator $\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}$ is equivalent to the time-limited Topelitz operator $\mathcal{H}_T$ in Section 1.2.2 and the corresponding eigenfunctions are known as PSWFs.
- As an another example, suppose $\mathbb{G} = \mathbb{Z}$, $\widehat{\mathbb{G}} = [-\frac{1}{2}, \frac{1}{2}]$ and let $\mathbb{A}_N = \{0, 1 \ldots, N-1\}$ (where $\tau = N$ in (26)), $\mathbb{B} = [-W, W]$ with $W \in (0, \frac{1}{2})$. In this case, the kernel $K_{\mathbb{B}}(n)$ becomes $K_{\mathbb{B}}[n] = \int_{-W}^{W} e^{j2\pi fn}\,\mathrm{d}\,f = \frac{\sin(2\pi Wn)}{\pi Wn}$. Then plugging in this form into (39) gives [92, Theorem 3.2] $\sum_\ell(\lambda_\ell(\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}))^2 = 2NW - O(\log(2NW))$. We note that in this case, the operator $\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}$ is equivalent to the $N \times N$ prolate matrix $\boldsymbol{B}_{N,W}$ with entries

$$\boldsymbol{B}_{N,W}[m, n] := \frac{\sin(2\pi W(m-n))}{\pi(m-n)} \tag{29}$$

  for all $m, n \in \{0, 1, \ldots, N-1\}$. The eigenvalues and eigenvectors of the matrix $\boldsymbol{B}_{N,W}$ are referred to as the DPSS eigenvalues and DPSS vectors, respectively.
- As a final example, we consider $\mathbb{G} = \mathbb{Z}_N$, $\widehat{\mathbb{G}} = \mathbb{Z}_N$ and the Fourier transform is the conventional DFT. Suppose $M, K \le N$. Let $\tau = M$ in (26) and

$$\mathbb{A}_M = \{0, 1, \ldots, M-1\}\,,\ \mathbb{B} = \{0, 1, \ldots, K-1\}\,, \tag{30}$$

  In this case, $\chi_k[n] = e^{j2\pi\frac{nk}{N}}$ and the kernel $K_{\mathbb{B}}[n]$ is $K_{\mathbb{B}}[n] = \sum_{k=0}^{K-1} e^{j2\pi\frac{nk}{N}} = e^{j\pi n\frac{K-1}{N}}\frac{\sin(\pi\frac{nK}{N})}{\sin(\pi\frac{n}{N})}$. Then plugging in this form into (39) gives [17, 87] $\sum_\ell(\lambda_\ell(\mathcal{O}_{\mathbb{A}_M,\mathbb{B}}))^2 = \frac{MK}{N} - O(\log(\frac{MK}{N}))$.

Through the above examples, one may wonder whether we can in general replace the second term in (27) by $O(\log(|\mathbb{A}_\tau||\mathbb{B}|))$ with a finer analysis of $\sum_\ell(\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}))^2$. We utilize a two-dimensional example to answer this question in the negative: Suppose $\mathbb{G} = \mathbb{Z}^2$, $\widehat{\mathbb{G}} = [-\frac{1}{2}, \frac{1}{2}]\times[-\frac{1}{2}, \frac{1}{2}]$ and let

$\mathbb{A} = \{0, 1 \ldots, N-1\} \times \{0, 1 \ldots, N-1\}, \mathbb{B} = [-W, W] \times [-W, W]$ with $W \in (0, \frac{1}{2})$. In this case, the kernel $K_{\mathbb{B}}[n_1, n_2]$ is

$$K_{\mathbb{B}}[n_1, n_2] = \int_{-W}^{W} \int_{-W}^{W} e^{j2\pi f_1 n_1} e^{j2\pi f_2 n_2} \, \mathrm{d}\, f_1 \, \mathrm{d}\, f_2 = \frac{\sin(2\pi W n_1)}{\pi W n_1} \frac{\sin(2\pi W n_2)}{\pi W n_2}.$$

The eigenvectors of the corresponding operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ are known as the two-dimensional DPSSs. For this case, we have $\sum_\ell (\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}))^2 = 4N^2W^2 - O(NW \log(NW))$. In other words, in this case, we can only improve the second term in (27) to $O(NW \log(NW))$ rather than $O(\log(4N^2W^2))$.

Remark 2. We note that the transition region in (28) depends on ϵ in the form of $\frac{1}{\epsilon(1-\epsilon)}$. A better understanding of the transition region requires further complicated analysis. In the literature, finer results on the transition region are known for several common cases:

- The results for the eigenvalue distribution of the continuous time-frequency localization operator (where $\mathbb{G} = \mathbb{R}$, $\widehat{\mathbb{G}} = \mathbb{R}$, $\mathbb{A}_T = [-\frac{T}{2}, \frac{T}{2}]$ and $\mathbb{B} = [-\frac{1}{2}, \frac{1}{2}]$) has a rich history. As one example, for any $\epsilon \in (0, 1)$, Landau and Widom [45] provided the following asymptotic result $\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; [\epsilon, 1]) = T + \left(\frac{1}{\pi^2} \log \frac{1-\epsilon}{\epsilon}\right) \log \frac{\pi T}{2} + o\left(\log \frac{\pi T}{2}\right)$. This asymptotic result ensures the $O(\log(\frac{1}{\epsilon}) \log(T))$ dependence on ϵ and time-frequency area T. Recently, Osipov [51] proved that $\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; [\epsilon, 1]) \leq T + C \log(T)^2 \log(1/\epsilon)$, where C is a constant. Israel [30] provided a non-asymptotic bound on the number of eigenvalues in the transition region. Fix $\eta \in (0, 1/2]$. Given $\epsilon \in (0, 1/2)$ and $T \geq 2$, then [30]

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}; (\epsilon, 1-\epsilon)) \leq 2C_\eta \left(\log\left(\frac{\log T}{\epsilon}\right)\right)^{1+\eta} \log\left(\frac{T}{\epsilon}\right), \tag{31}$$

  where C_η is a constant dependent on $\eta \in (0, \frac{1}{2}]$.
- The earliest result on the eigenvalue distribution of the discrete time-frequency localization operator (where $\mathbb{G} = \mathbb{Z}$, $\widehat{\mathbb{G}} = [-\frac{1}{2}, \frac{1}{2})$, $\mathbb{A}_N = \{0, 1, \ldots, N-1\}$ and $\mathbb{B} = [-W, W]$ with $W \in (0, \frac{1}{2})$) comes from Slepian [67], who showed that for any $b \in \mathbb{R}$, asymptotically the DPSS eigenvalue $\lambda_\ell(\mathcal{O}(\mathbb{A}_N, \mathbb{B})) \to \frac{1}{1+e^{\pi b}}$ as $N \to \infty$ if $\ell = \lfloor 2NW + \frac{b}{\pi} \log N \rfloor$. This implies the asymptotic result: $\mathcal{N}(\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}; (\epsilon, 1-\epsilon)) \sim \frac{2}{\pi^2} \log N \log\left(\frac{1}{\epsilon} - 1\right)$. Recently, by examining the difference between the operator $\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}$ and the one formed by a partial DFT matrix, we have shown [36,37] the following nonasymptotic result characterizing the $O(\log N \log \frac{1}{\epsilon})$ dependence:

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_N,\mathbb{B}}; (\epsilon, 1-\epsilon)) \leq \left(\frac{8}{\pi^2} \log(8N) + 12\right) \log\left(\frac{15}{\epsilon}\right). \tag{32}$$

  The right hand side is further improved to $\frac{2}{\pi^2} \log(4N) \left(\frac{4}{\epsilon(1-\epsilon)}\right)$ in [34].
- We [87] have also provided similar results for the eigenvalue distribution of discrete periodic time-frequency localization operator with sets $\mathbb{A}_M$ and $\mathbb{B}$ defined in (30):

(33)

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_M,\mathbb{B}}; (\epsilon, 1-\epsilon)) \leq \left(\frac{8}{\pi^2} \log(8N) + 12\right) \log\left(\frac{15}{\epsilon}\right) + 4 \max\left(\frac{-\log\left(\frac{\pi}{32}\left(\left(\frac{M}{N}\right)^2 - 1\right)\epsilon\right)}{\log\left(\frac{M}{N}\right)}, 0\right).$$

Remark 3. It is also of particular interest to have a finer result on the number of eigenvalues that is greater than $\frac{1}{2}$ since this together with the size of the transition region gives us a complete understanding of the eigenvalue distribution.

- Landau [42] establishes the number of PSWF eigenvalues that are greater than $\frac{1}{2}$ as follows

$$\lambda_{(\lfloor T \rfloor - 1)}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}) \geq \frac{1}{2} \geq \lambda_{(\lceil T \rceil)}(\mathcal{O}_{\mathbb{A}_T,\mathbb{B}}). \tag{34}$$

- We [92] provided a similar result for the DPSS eigenvalues.

In the following two subsections, we review some applications of Theorem 3.

4.2. **Application: Communications.** In [19], Franceschetti extended Landau's theorem [41] for simple time and frequency intervals to other time and frequency sets of complicated shapes. Lim and Franceschetti [47] related the number of degrees of freedom of the space of band-limited signals to the deterministic notions of capacity and entropy. Now we apply Theorem 3 to the effective dimensionality of the "band-limited signals" observed over a finite set $\mathbb{A}$ by utilizing the result in Section 2.2. To that end, we plug $\widehat{\phi}(\xi) = 1_{\mathbb{B}}(\xi) = \begin{cases} 1, & \xi \in \mathbb{B} \\ 0, & \xi \notin \mathbb{B} \end{cases}$, the indicator function on $\mathbb{B}$, into (13) and get the following set of band-limited functions observed only over $\mathbb{A}$:

$$\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi)) := \left\{ x \in L_2(\mathbb{A}) : x(g) = \int_{\mathbb{B}} \alpha(\xi) \chi_\xi(g) \, \mathrm{d}\, \xi, \int |\alpha(\xi)|^2 \leq 1, g \in \mathbb{A} \right\}.$$

When $\mathbb{A} \subset \mathbb{R}^2$ represents a subset of time and space, the number of degrees of freedom in the set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$ determines the total amount of information that can be transmitted in time and space by multiple-scattered electromagnetic waves [19]. Now we turn to compute the effective dimensionality of the general set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$. In this case, $|\widehat{\phi}(\xi)|^2 = 1_{\mathbb{B}}(\xi)$ and the corresponding operator $\mathcal{A}\mathcal{A}^*$ defined in (17) is equivalent to the time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ in (25). Now Proposition 1 implies that the effective dimensionality $\mathcal{N}(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi)), \epsilon)$ is equal to the number of eigenvalues of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ that are greater than ϵ, which is given by Theorem 3. In words, the effective dimensionality of the set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$ is essentially $|\mathbb{A}||\mathbb{B}|$, and is insensitive to the level ϵ (as illustrated in (31)-(33), in many cases, this dimensionality only has $\log(\frac{1}{\epsilon})$ dependence on ϵ).

4.3. **Application: Signal representation.** In addition to the eigenvalues of the time-frequency limiting operator $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$, the eigenfunctions of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ are also of significant importance, owing to their concentration in the time and frequency domains. To see this, let $u_\ell(g)$ be the ℓ-th eigenfunction of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$, corresponding to the ℓ-th eigenvalue $\lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}})$. Denoting the Fourier transform of $u_\ell(g)$ by $\widehat{u}_\ell(\xi)$, we have

$$\begin{aligned} \int_{\mathbb{B}} |\widehat{u}_\ell(\xi)|^2 \, \mathrm{d}\, \xi &= \langle \mathcal{T}_{\mathbb{B}}\mathcal{F}u_\ell, \mathcal{T}_{\mathbb{B}}\mathcal{F}u_\ell \rangle = \left\langle \mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F}u_\ell, u_\ell \right\rangle = \left\langle \mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F}\mathcal{T}_{\mathbb{A}}u_\ell, \mathcal{T}_{\mathbb{A}}u_\ell \right\rangle \\ &= \left\langle \mathcal{T}_{\mathbb{A}}\mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F}\mathcal{T}_{\mathbb{A}}u_\ell, \mathcal{T}_{\mathbb{A}}u_\ell \right\rangle = \lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}) \|\mathcal{T}_{\mathbb{A}}u_\ell\|^2, \end{aligned} \tag{35}$$

where the third equality follows because $u_\ell(g)$ is a time-limited signal (i.e., $\mathcal{T}_{\mathbb{A}}(u_\ell) = u_\ell$), and the last equality utilizes $\mathcal{T}_{\mathbb{A}}\mathcal{F}^{-1}\mathcal{T}_{\mathbb{B}}\mathcal{F}\mathcal{T}_{\mathbb{A}}u_\ell = \mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}u_\ell = \lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}})u_\ell$. In words, (35) states that the eigenfunctions u_ℓ have a proportion $\lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{A}_{\mathbb{A}})$ of energy within the band $\mathbb{B}$, implying that even though the eigenfunctions are not exactly band-limited, their Fourier transform is mostly concentrated in the band $\mathbb{B}$ when $\lambda_\ell(\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{A}_{\mathbb{A}})$ is close to 1. Thus, the first $\approx |\mathbb{A}||\mathbb{B}|$ eigenfunctions can be utilized as window functions for spectral estimation, and as a highly efficient basis for representing band-limited signals that are observed over a finite set $\mathbb{A}$.

Recall that $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$ (defined in (13)) consists of band-limited signals observed over a finite set $\mathbb{A}$. Applying Proposition 1, we compute the n-width of the set $\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$ as follows:

$$d_n(\mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))) = \sqrt{\lambda_n(\mathcal{A}\mathcal{A}^*)} = \sqrt{\lambda_n(\mathcal{O}_{\mathbb{A},\mathbb{B}})}.$$

By the definition of (15), we know for any $x(g) \in \mathbb{W}(\mathbb{A}, 1_{\mathbb{B}}(\xi))$,

$$\int_{\mathbb{A}} |x(g) - (\mathrm{P}_{\mathbb{U}_n} x)(g)|^2 \, \mathrm{d}\, g \leq \sqrt{\lambda_n(\mathcal{O}_{\mathbb{A},\mathbb{B}})},$$

where $\mathbb{U}_n$ is the subspace spanned by the first n eigenvectors of $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$, i.e.,

$$\mathbb{U}_n := \operatorname{span}\{u_0(g), u_1(g), \ldots, u_{n-1}(g)\}. \tag{36}$$

Now we utilize Theorem 3 to conclude that the representation residual $\sqrt{\lambda_n(\mathcal{O}_{\mathbb{A},\mathbb{B}})}$ is very small when n is chosen slightly larger than $|\mathbb{A}||\mathbb{B}|$.

We now investigate the basis $\mathbb{U}_n$ for representing time-limited version of characters $\chi_\xi(g)$ and band-limited signals.

4.3.1. *Approximation quality for time-limited characters $\chi_\xi(g)$.* We first restrict our focus to the simplest possible "band-limited signals" that are observed over a finite period: pure characters $\chi_\xi(g)$ when g is limited to $\mathbb{A}$. Without knowing the exact frequency ξ in advance, we attempt to find an efficient low-dimensional basis for capturing the energy in any signal $\chi_\xi(g)$. To that end, we let $\mathbb{M}_n \subset L_2(\mathbb{A})$ denote an n-dimensional subspace of $L_2(\mathbb{A})$. We would like to minimize

$$\int_{\mathbb{B}} \|\chi_\xi - \mathrm{P}_{\mathbb{M}_n}\chi_\xi\|^2_{L_2(\mathbb{A})}\,\mathrm{d}\,\xi. \tag{37}$$

The following result establishes the degree of approximation accuracy in a mean-squared error (MSE) sense provided by the subspace $\mathbb{U}_n$ for representing the "time-limited" version of characters $\chi_\xi(g)$ (where g is limited to $\mathbb{A}_\tau$).

Theorem 4. *For any $n \in \mathbb{Z}^+$, the optimal n-dimensional subspace which minimizes (37) is $\mathbb{U}_n$. Furthermore, with this choice of subspace, we have*

$$\frac{1}{|\mathbb{B}|}\int_{\mathbb{B}} \frac{\|\chi_\xi - \mathrm{P}_{\mathbb{U}_n}\chi_\xi\|^2_{L_2(\mathbb{A})}}{\|\chi_\xi\|^2_{L_2(\mathbb{A})}}\,\mathrm{d}\,\xi = 1 - \frac{\sum_{\ell=0}^{n-1}\lambda_\ell(\mathcal{O}_{\mathbb{A},\mathbb{B}})}{|\mathbb{A}||\mathbb{B}|},$$

where $|\cdot|$ denotes the Haar measure.

The proof of Theorem 4 is given in Appendix C. Combined with Theorem 3, Theorem 4 implies that by choosing $n \approx |\mathbb{A}||\mathbb{B}|$, on average the subspace spanned by the first n eigenfunctions of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ is expected to accurately represent time-limited characters within the band of interest. We note that the representation guarantee for time-limited characters $\{\mathcal{T}_{\mathbb{A}}\chi_\xi, \xi \in \mathbb{B}\}$ can also be used for most band-limited signals that are observed over a finite set $\mathbb{A}$. To see this, suppose $x(g)$ is a band-limited function which can be represented as

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

An immediate consequence of the above equation is that one can view $\{\mathcal{T}_{\mathbb{A}}\chi_\xi, \xi \in \mathbb{B}\}$ as the atoms for building $\mathcal{T}_{\mathbb{A}}x$:

$$\mathcal{T}_{\mathbb{A}}x = \int_{\mathbb{B}} \widehat{x}(\xi)\mathcal{T}_{\mathbb{A}}\chi_\xi\,\mathrm{d}\,\xi.$$

4.3.2. *Approximation quality for random band-limited signals.* We can also approach the representation ability of the subspace $\mathbb{U}_n$ (defined in (36)) from a probabilistic perspective by generalizing [12, Theorem 4.1].

Theorem 5. *Let $x(g) = \chi_\xi(g), g \in \mathbb{A}_\tau$ be a random function where ξ is a random variable with uniform distribution on $\mathbb{B}$. Then we have*

$$\frac{\mathrm{E}\left[\|x - \mathrm{P}_{\mathbb{U}_n}x\|^2_{L_2(\mathbb{A})}\right]}{\mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A})}\right]} = 1 - \frac{\sum_{\ell=0}^{n-1}\lambda_\ell(\mathcal{O}_{\mathbb{A},\mathbb{B}})}{|\mathbb{A}||\mathbb{B}|}.$$

The proof of Theorem 5 is given in Appendix D. With this result, we show that in a certain sense, most band-limited signals, when time-limited, are well-approximated by a signal within the subspace $\mathbb{U}_n$. In particular, the following result which generalizes [12, Theorem 4.1] establishes that band-limited random processes, when time-limited, are in expectation well-approximated.

Corollary 2. *Let $x(g), g \in \mathbb{G}$ be a zero-mean wide sense stationary random process over the group $\mathbb{G}$ with power spectrum*

$$P_x(\xi) = \begin{cases} \frac{1}{|\mathbb{B}|}, & \xi \in \mathbb{B}, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose we only observe x over the set $\mathbb{A}_\tau$. Then we have

$$\frac{\mathrm{E}\left[\|x - \mathrm{P}_{\mathbb{U}_n} x\|^2_{L_2(\mathbb{A})}\right]}{\mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A})}\right]} = 1 - \frac{\sum_{\ell=0}^{n-1} \lambda_\ell(\mathcal{O}_{\mathbb{A},\mathbb{B}})}{|\mathbb{A}||\mathbb{B}|}.$$

As in our discussion following Theorem 3, the term $1 - \frac{\sum_{\ell=0}^{n-1} \lambda_\ell(\mathcal{O}_{\mathbb{A},\mathbb{B}})}{|\mathbb{A}||\mathbb{B}|}$ appearing in Theorem 5 and Corollary 2 can be very small when we choose n slightly larger than $|\mathbb{A}||\mathbb{B}|$. This suggests that in a probabilistic sense, most band-limited functions, when time-limited, will be well-approximated by a small number of eigenfunctions of the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$.

4.4. **Applications in the common time and frequency domains.** We now review several applications involving the time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ in the common time and frequency domains, where the eigenfunctions correspond to DPSSs, PSWFs, and PDPSSs.

It follows from (35) that, among all the functions that are time-limited to the set $\mathbb{A}$, the first eigenfunction $u_0(g)$ is maximally concentrated in the subset $\mathbb{B}$ of the frequency domain. Motivated by this result, the first DPSS vector is utilized as a filter for super-resolution [18]. In [73], the first $\approx 2NW$ DPSS vectors are utilized as window functions (a.k.a. tapers) for spectral estimation. The multitaper method [73] averages the tapered estimates with the DPSS vectors, and has been used in a variety of scientific applications including statistical signal analysis [10], geophysics and cosmology [11]. By exploiting the fact that the number of DPSS eigenvalues in the transition region grows like $O(\log N \log \frac{1}{\epsilon})$ as in (32), the very recent work [35] provided nonasymptotic bounds on some statistical properties of the multitaper spectral estimate as well as a fast algorithm for evaluating the estimate.

By exploiting the concentration behavior of the PSWFs in the time and frequency domains (where $\mathbb{G} = \mathbb{R}$ and $\widehat{\mathbb{G}} = \mathbb{R}$), Xiao et al. [79] utilized the PSWFs to numerically construct quadratures, interpolation and differentiation formulae for band-limited functions. Gosse [21] constructed a PSWF dictionary consisting of the first few PSWFs for recovering smooth functions from random samples. The connection between time-frequency localization of multiband signals and sampling theory for such signals was investigated in [31]. In [62, 63], the authors also considered a PSWF dictionary for reconstruction of electroencephalography (EEG) signals and time-limited signals that are also nearly band-limited from nonuniform samples. Chen and Vaidyanathan [7] utilized the PSWFs to represent the clutter subspace (and hence mitigate the clutter), facilitating space-time adaptive processing for multiple-input multiple-output (MIMO) radar systems; see also [16, 81].

DPSSs, the discrete counterpart of PSWFs, also have proved to be useful in numerous signal processing applications since they provide a highly efficient basis for representing sampled band-limited signals. DPSSs can be utilized to find the minimum energy, infinite-length band-limited sequence that extrapolates a given finite vector of samples [67]. In [84, 85], Zemen et al. expressed the time-varying subcarrier coefficients in a DPSS basis for estimating time-varying channels in wireless communication systems. A similar idea is also utilized for channel estimation in Orthogonal Frequency Division Multiplexing (OFDM) systems [70], for receiver antenna selection [57], etc.

The modulated DPSSs can also be useful for mitigating wall clutter and detecting targets behind the wall in through-the-wall radar imaging [90, 91], and for interference cancellation in a wideband compressive radio receiver (WCRR) architecture [13]. The performance (such as the detection probability) of the DPSS basis (and other similar bases corresponding to the time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$) for identifying unresolved targets was recently analyzed in [5]. 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 [12, 92]. Sejdić et al. [61] proposed one such dictionary to provide a sparse representation for fading channels and improve channel estimation accuracy. The multiband modulated DPSS dictionaries have been utilized for the recovery of sampled multiband signals from random measurements [12], and for the recovery of physiological signals from compressive measurements [60]. Such dictionaries are also utilized for cancelling wall clutter [1].

The periodic DPSSs (PDPSSs, where $\mathbb{G} = \mathbb{Z}_N$ and $\widehat{\mathbb{G}} = \mathbb{Z}_N$) are the finite-length vectors whose discrete Fourier transform (DFT) is most concentrated in a given bandwidth (as appearing in (30)). The PDPSSs have been utilized for extrapolation and spectral estimation of periodic discrete-time signals [32], for limited-angle reconstruction in tomography [25], for Fourier extension [50], and in [29], the bandpass PDPSSs were used as a numerical approximation to the bandpass PSWFs for studying synchrony in sampled EEG signals.

Finally, the eigenvalue concentration behavior in Theorem 3 can also be exploited for solving a linear system involving the Toeplitz operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$: $y = \mathcal{O}_{\mathbb{A},\mathbb{B}}x$. Since the operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ has a mass of eigenvalues that are very close to 0, the system is often solved by using the rank-K pseudoinverse of $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ where $K \approx |\mathbb{A}||\mathbb{B}|$. In the case where the Toeplitz operator is the prolate matrix $\boldsymbol{B}_{N,W}$ defined in (29), its truncated pseudoinverse is well approximated as the sum of $\boldsymbol{B}^*_{N,W}$ (which is equal to $\boldsymbol{B}_{N,W}$) and a low-rank matrix [37, 50] since most of the eigenvalues of $\boldsymbol{B}^*_{N,W}$ are very close to either 1 or 0. By utilizing the fact that $\boldsymbol{B}_{N,W}$ is a Toeplitz matrix and $\boldsymbol{B}_{N,W}\boldsymbol{x}$ has a fast implementation via the FFT, an efficient method for solving the system $\boldsymbol{y} = \boldsymbol{B}_{N,W}\boldsymbol{x}$ can be developed; such a method has been utilized for linear prediction of band-limited signals based on past samples and the Fourier extension [37, 50].

4.5. **Questions.** Inspired by the applications listed above, we raise several questions concerning Theorems 3 and 4. Following from the two remarks after Theorem 3, two natural questions are:

Question 3. *Can we improve the second term in (27)? Furthermore, what nonasymptotic result (like (31) for the PSWF eigenvalues and (32) for the DPSS eigenvalues) can we obtain for the number of eigenvalues of the Toeplitz operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$ within the transition region $(\epsilon, 1-\epsilon)$?*

Question 4. *Can we extend (34) to the general time-frequency limiting operator $\mathcal{O}_{\mathbb{A},\mathbb{B}}$?*

Another related important question concerns how accurately the subspace spanned by the first n eigenfunctions of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ can represent each individual time-limited character $\mathcal{T}_{\mathbb{A}}\chi_\xi$ with $\xi \in \mathbb{B}$. Theorem 4 ensures that accuracy is guaranteed in the MSE sense if one chooses $n \approx |\mathbb{A}||\mathbb{B}|$ such that the sum of the remaining eigenvalues of $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$ is small. We suspect that a uniform guarantee for each $\mathcal{T}_{\mathbb{A}}\chi_\xi$ can also be obtained since the derivative of $\|\chi_\xi\|^2_{L_2(\mathbb{A})}$ is bounded, given a finer result concerning the eigenvalue distribution for $\mathcal{T}_{\mathbb{A}}\mathcal{B}_{\mathbb{B}}\mathcal{T}_{\mathbb{A}}$. Using the approach utilized in [88, 89] with a theorem of Bernstein for trigonometric polynomials [59], we can have an approximation guarantee for the DPSS basis in representing each complex exponential $\boldsymbol{e}_f := \begin{bmatrix} e^{j2\pi f 0} & \cdots & e^{j2\pi f(N-1)} \end{bmatrix}^{\mathrm{T}}$ with frequency f inside a band of interest; this provides a non-asymptotic guarantee which improves upon our previous work [92].

Theorem 6. *(Representation guarantee for pure sinusoids with DPSSs) Let $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$ be given. Also let $[\boldsymbol{S}]_K$ be an $N \times K$ matrix consisting of the first K DPSS vectors. Then for any*

$\epsilon \in (0, \frac{1}{2})$, *the orthobasis* $[\boldsymbol{S}]_K$ *satisfies*

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

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

$$K = 2NW + O\left(\log(N) \log\left(\frac{1}{\epsilon^2}\right)\right).$$

Question 5. *More generally, what uniform guarantee can we have for each time-limited character* $\mathcal{T}_{\mathbb{A}} \chi_\xi$ *in the subspace spanned by the first* n *eigenfunctions of* $\mathcal{T}_{\mathbb{A}} \mathcal{B}_{\mathbb{B}} \mathcal{T}_{\mathbb{A}}$*?*

Acknowledgements

This work was supported by NSF grants CCF-1409261, CCF-1704204, and CCF-2008460. We gratefully acknowledge Justin Romberg for pointing out a connection between subspace approximations and Toeplitz operators, and Mark Davenport and Santhosh Karnik for valuable discussions.

Appendix A. Proof of Proposition 1

Proof of Proposition 1. We have

$$\begin{aligned}
d_n(\mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))) &= \inf_{\mathbb{M}_n} \sup_{x \in \mathbb{W}(\mathbb{A}, \widehat{\phi}(\xi))} \inf_{y \in \mathbb{M}_n} \|x - y\|_{L_2(\mathbb{A})} \\
&= \inf_{\mathbb{M}_n} \sup_{\|\alpha\| \leq 1} \|\mathcal{A}\alpha - \mathrm{P}_{\mathbb{M}_n}\| \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \sup_{\|\alpha\| \leq 1} \frac{|\langle \mathcal{A}\alpha, z \rangle|}{\|z\|} \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \sup_{\|\alpha\| \leq 1} \frac{|\langle \alpha, \mathcal{A}^* z \rangle|}{\|z\|} \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \frac{\|\mathcal{A}^* z\|}{\|z\|} \\
&= \inf_{\mathbb{M}_n} \sup_{z \perp \mathbb{M}_n} \frac{\sqrt{\langle \mathcal{A}\mathcal{A}^* z, z \rangle}}{\|z\|} \\
&= \sqrt{\lambda_n},
\end{aligned}$$

where the last line follows from the Weyl-Courant minimax theorem. □

Appendix B. Proof of Theorem 3

Proof of Theorem 3. We first note that $\chi_\xi(0) = 1$ for all $\xi \in \widehat{\mathbb{G}}$. Thus, we have

$$\sum_\ell \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau, \mathbb{B}}) = \int_{\mathbb{A}_\tau} K_{\mathbb{B}}(0) \, \mathrm{d}\, h = |\mathbb{A}_\tau| \int_{\mathbb{B}} \chi_\xi(0) \, \mathrm{d}\, \xi = |\mathbb{A}_\tau||\mathbb{B}|, \tag{38}$$

where the first equality follows because $\mathcal{O}_{\mathbb{A}_\tau, \mathbb{B}}$ is a trace class operator [14]. We write the operator $(\mathcal{T}_{\mathbb{A}_\tau} \mathcal{B}_{\mathbb{B}} \mathcal{T}_{\mathbb{A}_\tau})^2$ as

$$\begin{aligned}
(\mathcal{T}_{\mathbb{A}_\tau} \mathcal{B}_{\mathbb{B}} \mathcal{T}_{\mathbb{A}_\tau} \mathcal{B}_{\mathbb{B}} \mathcal{T}_{\mathbb{A}_\tau} x)(g) &= \int_{\mathbb{A}_\tau} K_{\mathbb{B}}(\widetilde{h}^{-1} \circ g) \left(\int_{\mathbb{A}_\tau} K_{\mathbb{B}}(h^{-1} \circ \widetilde{h}) x(h) \, \mathrm{d}\, h \right) \mathrm{d}\, \widetilde{h} \\
&= \int_{\mathbb{A}_\tau} \left(\int_{\mathbb{A}_\tau} K_{\mathbb{B}}(\widetilde{h}^{-1} \circ g) K_{\mathbb{B}}(h^{-1} \circ \widetilde{h}) \, \mathrm{d}\, \widetilde{h} \right) x(h) \, \mathrm{d}\, h.
\end{aligned}$$

Thus,

$$\sum_{\ell} \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \int_{\mathbb{A}_\tau}\int_{\mathbb{A}_\tau} K_{\mathbb{B}}(\widetilde{h}^{-1}\circ h)K_{\mathbb{B}}(h^{-1}\circ \widetilde{h})\,\mathrm{d}\,\widetilde{h}\,\mathrm{d}\,h = \int_{\mathbb{A}_\tau}\int_{\mathbb{A}_\tau} \left|K_{\mathbb{B}}(h^{-1}\circ \widetilde{h})\right|^2 \mathrm{d}\,\widetilde{h}\,\mathrm{d}\,h,$$

where we use the fact that $K_{\mathbb{B}}(h^{-1}\circ g) = \int_{\mathbb{B}} \chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi = (\int_{\mathbb{B}} \chi_\xi(g^{-1}\circ h)\,\mathrm{d}\,\xi)^*$ since $\chi_\xi(-g) = \chi_\xi^*(g)$. Applying the change of variable $\widetilde{h} = h\circ \overline{h}$, we obtain

$$\sum_{\ell} \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \int_{\mathbb{A}_\tau}\int_{\mathbb{A}_\tau - h} \left|K_{\mathbb{B}}(\overline{h})\right|^2 \mathrm{d}\,\overline{h}\,\mathrm{d}\,h = \int_{\mathbb{A}_\tau} \kappa_{\mathbb{A}_\tau,\mathbb{B}}(h)\,\mathrm{d}\,h, \tag{39}$$

where $\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h) = \int_{\mathbb{A}_\tau - h} \left|K_{\mathbb{B}}(\overline{h})\right|^2 \mathrm{d}\,\overline{h} \geq 0$. The function $\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h)$ is dominated as

$$\kappa_{\mathbb{A}_\tau,\mathbb{B}}(h) \leq \int_{\mathbb{G}} \left|K_{\mathbb{B}}(\overline{h})\right|^2 \mathrm{d}\,\overline{h} = \int_{\mathbb{G}} \left|\int_{\mathbb{B}} \chi_\xi(\overline{h})\,\mathrm{d}\,\xi\right|^2 \mathrm{d}\,\overline{h} = \int_{\widehat{\mathbb{G}}} 1_{\xi\in\mathbb{B}}\,\mathrm{d}\,\xi = |\mathbb{B}|,$$

where we use Parseval's theorem by viewing $\int_{\mathbb{B}} \chi_\xi(\overline{h})\,\mathrm{d}\,\xi$ as the inverse Fourier transform of a window function supported on $\mathbb{B}$, i.e. $1_{\xi\in\mathbb{B}} = \begin{cases} 1, & \xi\in\mathbb{B}, \\ 0, & \text{otherwise}, \end{cases}$.

On the other hand, we have $\lim_{\tau\to\infty} \kappa_{\mathbb{A}_\tau,\mathbb{B}}(h) = \int_{\mathbb{G}} \left|\int_{\mathbb{B}} \chi_\xi(\overline{h})\,\mathrm{d}\,\xi\right|^2 \mathrm{d}\,\overline{h} = |\mathbb{B}|$ for all $h\in\mathbb{G}$. It follows that $\lim_{\tau\to\infty}\sum_\ell \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \int_{\mathbb{A}_\tau} |\mathbb{B}|\,\mathrm{d}\,h = |\mathbb{A}_\tau||\mathbb{B}|$. Thus, we have

$$\sum_{\ell} \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{A}_\tau||\mathbb{B}| - o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{40}$$

Subtracting (40) from (38) gives

$$\sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\,(1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{41}$$

Utilizing the fact that $0\leq \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq 1$, we have

$$\epsilon(1-\epsilon)\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};(\epsilon,1-\epsilon)) \leq \sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\,(1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})) = o(|\mathbb{A}_\tau||\mathbb{B}|).$$

On the other hand, (41) also implies that

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})<1-\epsilon} \epsilon\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) < \sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\,(1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{42}$$

Plugging this term into (38) gives

$$\begin{aligned} |\mathbb{A}_\tau||\mathbb{B}| = \sum_{\ell} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) &= \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) + \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})<1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \\ &= \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) + o(|\mathbb{A}_\tau||\mathbb{B}|). \end{aligned}$$

Similarly, plugging (42) into (40) gives

$$|\mathbb{A}_\tau||\mathbb{B}| = \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) + o(|\mathbb{A}_\tau||\mathbb{B}|).$$

Combining the above two equations and the fact that $\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq 1$, we have

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) - \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = o(|\mathbb{A}_\tau||\mathbb{B}|). \tag{43}$$

On one hand, combining (43) with

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) - \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} 1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})$$

gives

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) - \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} 1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \geq o(|\mathbb{A}_\tau||\mathbb{B}|),$$

which further implies

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) \geq |\mathbb{A}_\tau||\mathbb{B}| - o(|\mathbb{A}_\tau||\mathbb{B}|).$$

On the other hand, using (43) and

$$\sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) - \lambda_\ell^2(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \geq (1-\epsilon) \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\leq 1-\epsilon} 1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}),$$

we also have

$$\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) - \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} \lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = \sum_{\ell:\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})\geq 1-\epsilon} 1-\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) \leq o(|\mathbb{A}_\tau||\mathbb{B}|),$$

which further implies $\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1]) \leq |\mathbb{A}_\tau||\mathbb{B}|+o(|\mathbb{A}_\tau||\mathbb{B}|)$. Thus, $\lim_{\tau\to\infty} \frac{\mathcal{N}(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}};[1-\epsilon,1])}{|\mathbb{A}_\tau|} = |\mathbb{B}|$. □

Appendix C. Proof of Theorem 4

Proof of Theorem 4. We first recall the eigendecompostion of $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}} = \sum_{\ell\geq 0} \lambda_\ell u_\ell u_\ell^*$, where λ_ℓ is short for $\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}})$. Utilizing the fact that $u_\ell, \ell = 0,1,\ldots$ is a complete orthonormal basis for $L_2(\mathbb{A}_\tau)$, we rewrite the function in (37):

$$\begin{aligned}
\|\chi_\xi(g) - \mathrm{P}_{\mathbb{M}_n}\chi_\xi(g)\|_{L_2(\mathbb{A}_\tau)}^2 &= \sum_\ell \left|\langle (I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(g), u_\ell(g)\rangle_{L_2(\mathbb{A}_\tau)}\right|^2 \\
&= \sum_\ell \left\langle \left\langle (I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(g)\chi_\xi^*(h), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)} \\
&= \sum_\ell \left\langle \left\langle (I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(h^{-1}\circ g), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)}
\end{aligned}$$

where the second equality utilized the fact that $\mathrm{P}_{\mathbb{M}_n}$ is the orthogonal projector onto the subspace $\mathbb{M}_n$, and $\sum_\ell \left\langle \left\langle (I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(g)\chi_\xi^*(h), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(h)\right\rangle_{L_2(\mathbb{A}_\tau)}$ is equivalent to the trace of $(I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(g)\chi_\xi^*(h)$. Plugging this equation into (37) gives

$$\begin{aligned}
\int_{\mathbb{B}} \|\chi_\xi(g) - \mathrm{P}_{\mathbb{M}_n}\chi_\xi(g)\|_{L_2(\mathbb{A}_\tau)}^2 \,\mathrm{d}\,\xi &= \int_{\mathbb{B}} \sum_\ell \left\langle \left\langle (I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(\theta^{-1}\circ g), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)} \mathrm{d}\,\xi \\
&= \sum_\ell \int_{\mathbb{B}} \left\langle \left\langle (I-\mathrm{P}_{\mathbb{M}_n})\chi_\xi(\theta^{-1}\circ g), u_\ell^*(h)\right\rangle_{L_2(\mathbb{A}_\tau)}, u_\ell(g)\right\rangle_{L_2(\mathbb{A}_\tau)} \mathrm{d}\,\xi \\
&= \sum_\ell \langle (I-\mathrm{P}_{\mathbb{M}_n})\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}u_\ell, u_\ell\rangle_{L_2(\mathbb{A}_\tau)} = \sum_\ell \lambda_\ell \langle (I-\mathrm{P}_{\mathbb{M}_n})u_\ell, u_\ell\rangle_{L_2(\mathbb{A}_\tau)}
\end{aligned}$$

where the second line follows from monotone convergence theorem (since each term inside the summation is nonnegative). Thus, we conclude that the optimal n-dimensional subspace which

minimizes the last term in the above equation is $\mathbb{U}_n$ (which is spanned by the first n eigenfunctions). With this choice of subspace and (27) that $\sum_\ell \lambda_\ell = |\mathbb{A}_\tau||\mathbb{B}|$, we have

$$\int_{\mathbb{B}} \|\chi_\xi(g) - \mathrm{P}_{\mathbb{U}_n}\chi_\xi(g)\|^2_{L_2(\mathbb{A}_\tau)}\,\mathrm{d}\,\xi = \sum_{\ell\geq n}\lambda_\ell = |\mathbb{A}_\tau||\mathbb{B}| - \sum_{\ell=0}^{n-1}\lambda_\ell$$

The proof is completed by noting that $\|\chi_\xi(g)\|^2_{L_2(\mathbb{A}_\tau)} = |\mathbb{A}_\tau|$ for any $\xi \in \mathbb{B}$. □

Appendix D. Proof of Theorem 5

Proof of Theorem 5. First let ν be a random variable with uniform distribution on $[0, 2\pi)$. We define the random vector

$$r(g) = r(g;\xi,\nu) = \chi_\xi(g)e^{j\nu},$$

where the term $e^{j\nu}$ acts as a phase randomizer and ensures that r is zero-mean:

$$\mathrm{E}\left[r(g)\right] = \frac{1}{|\mathbb{B}|2\pi}\int_{\mathbb{B}}\chi_\xi(g)e^{j\nu}\,\mathrm{d}\,\xi\,\mathrm{d}\,\nu = \frac{1}{|\mathbb{B}|2\pi}\int_{\mathbb{B}}\chi_\xi(g)\,\mathrm{d}\,\xi\int_0^{2\pi}e^{j\nu}\,\mathrm{d}\,\nu = 0$$

for all $g \in \mathbb{A}_\tau$.

Now we compute the autocorrelation R of the random variable r as

$$\begin{aligned} R(g,h) &= \mathrm{E}\left[r(g)r^*(h)\right] = \mathrm{E}\left[\left(\chi_\xi(g)e^{j\nu}\right)\left(\chi^*_\xi(h)e^{-j\nu}\right)\right] = \mathrm{E}\left[\chi_\xi(h^{-1}\circ g)\right] \\ &= \frac{1}{|\mathbb{B}|}\int_{\mathbb{B}}\chi_\xi(h^{-1}\circ g)\,\mathrm{d}\,\xi = \frac{1}{|\mathbb{B}|}K_{\mathbb{B}}(h^{-1}\circ g) \end{aligned} \tag{44}$$

for all $h, g \in \mathbb{A}_\tau$. Here $K_{\mathbb{B}}$ is defined in (24). Note that $K_{\mathbb{B}}(h^{-1}\circ g)$ with $h, g \in \mathbb{A}_\tau$ is the kernel of the Toeplitz operator $\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}$. Now it follows from the Karhunen-Loève (KL) transfrom [71] that

$$\mathrm{E}\left[\|r - \mathrm{P}_{\mathbb{U}_n}r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \frac{1}{|\mathbb{B}|}\sum_{\ell\geq n}\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}) = |\mathbb{B}| - \sum_{\ell=0}^{n-1}\lambda_\ell(\mathcal{O}_{\mathbb{A}_\tau,\mathbb{B}}).$$

We then compute the expectation for the energy of r as

$$\mathrm{E}\left[\|r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \frac{1}{|\mathbb{B}|}\frac{1}{2\pi}\int_{\mathbb{B}}|\chi_\xi(g)e^{j\nu}|^2\,\mathrm{d}\,\xi\,\mathrm{d}\,\nu = |\mathbb{A}_\tau|.$$

The proof is completed by noting that $\mathrm{E}\left[\|r - \mathrm{P}_{\mathbb{U}_n}r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \mathrm{E}\left[\|x - \mathrm{P}_{\mathbb{U}_n}x\|^2_{L_2(\mathbb{A}_\tau)}\right]$ and $\mathrm{E}\left[\|r\|^2_{L_2(\mathbb{A}_\tau)}\right] = \mathrm{E}\left[\|x\|^2_{L_2(\mathbb{A}_\tau)}\right]$. □

References

[1] F. Ahmad, Q. Jiang, and M. Amin. Wall clutter mitigation using discrete prolate spheroidal sequences for sparse reconstruction of indoor stationary scenes. *IEEE Trans. Geosci. Remote Sens.*, 53(3):1549–1557, 2015.

[2] A. Araujo, B. Negrevergne, Y. Chevaleyre, and J. Atif. On lipschitz regularization of convolutional layers using toeplitz matrix theory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6661–6669, 2021.

[3] F. Avram. On bilinear forms in gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields*, 79(1):37–45, 1988.

[4] J. Bogoya, A. Böttcher, S. Grudsky, and E. Maximenko. Maximum norm versions of the Szegő and Avram-Parter theorems for Toeplitz matrices. *J. Approx. Theory*, 196(0):79 – 100, 2015.

[5] J. Bosse and O. Rabaste. Subspace rejection for matching pursuit in the presence of unresolved targets. *IEEE Transactions on Signal Processing*, 66(8):1997–2010, 2018.

[6] C. Carathéodory. Über den variabilitätsbereich der fourierschen konstanten von positiven harmonischen funktionen. *Rendiconti Del Circolo Matematico di Palermo (1884-1940)*, 32(1):193–217, 1911.

[7] C.-Y. Chen and P. P. Vaidyanathan. MIMO radar space–time adaptive processing using prolate spheroidal wave functions. *IEEE Trans. Signal Process.*, 56(2):623–635, 2008.

[8] G. S. Chirikjian and A. B. Kyatkin. *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition*. Courier Dover Publications, 2016.
[9] J. B. Conway. *A course in functional analysis*, volume 96. Springer, 2019.
[10] D. D. Cox. Spectral analysis for physical applications: Multitaper and conventional univariate techniques, 1996.
[11] F. Dahlen and F. J. Simons. Spectral estimation on a sphere in geophysics and cosmology. *Geophys. J. Int.*, 174(3):774–807, 2008.
[12] M. Davenport and M. Wakin. Compressive sensing of analog signals using discrete prolate spheroidal sequences. *Appl. Comput. Harmon. Anal.*, 33(3):438–472, 2012.
[13] M. A. Davenport, S. R. Schnelle, J. Slavinsky, R. G. Baraniuk, M. B. Wakin, and P. T. Boufounos. A wideband compressive radio receiver. In *Military Communications Conf. (MILCOM)*, pages 1193–1198. IEEE, 2010.
[14] A. Deitmar and G. van Dijk. Trace class groups. *arXiv preprint arXiv:1501.02375*, 2015.
[15] G. T. Di Francia. Degrees of freedom of an image. *J. Opt. Soc. Am.*, 59(7):799–804, 1969.
[16] W. Du, G. Liao, and Z. Yang. Robust space time processing based on bi-iterative scheme of secondary data selection and PSWF method. *Digital Signal Process.*, 52:64–71, 2016.
[17] A. Edelman, P. McCorquodale, and S. Toledo. The future fast Fourier transform? *SIAM J. Sci. Comput.*, 20(3):1094–1114, 1998.
[18] A. Eftekhari and M. B. Wakin. Greed is super: A fast algorithm for super-resolution. *arXiv preprint arXiv:1511.03385*, 2015.
[19] M. Franceschetti. On Landau's eigenvalue theorem and information cut-sets. *IEEE Trans. Inform. Theory*, 61(9):5042–5051, 2015.
[20] H. Gazzah, P. A. Regalia, and J.-P. Delmas. Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification. *IEEE Trans. Inf. Theory*, 47(3):1243–1251, 2001.
[21] L. Gosse. Compressed sensing with preconditioning for sparse recovery with subsampled matrices of Slepian prolate functions. *Annali Dell'Universita'Di Ferrara*, 59(1):81–116, 2013.
[22] R. Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Trans. Inf. Theory*, 18(6):725–730, 1972.
[23] R. M. Gray. Toeplitz and circulant matrices: A review. *Commun. Inf. Theory*, 2(3):155–239, 2005.
[24] U. Grenander and G. Szegő. *Toeplitz Forms and Their Applications*, volume 321. Univ of California Press, 1958.
[25] F. A. Grünbaum. Eigenvectors of a Toeplitz matrix: Discrete version of the prolate spheroidal wave functions. *SIAM J. Algeb. Discrete Methods*, 2(2):136–141, 1981.
[26] J. Gutiérrez-Gutiérrez and P. M. Crespo. Asymptotically equivalent sequences of matrices and hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems. *IEEE Trans. Inf. Theory*, 54(12):5671–5680, 2008.
[27] I. Hirschman, D. Liang, and E. Wilson. Szegő limit theorems for toeplitz operators on compact homogeneous spaces. *Trans. Amer. Math. Soc.*, 270(2):351–376, 1982.
[28] J. A. Hogan and J. D. Lakey. *Duration and Bandwidth Limiting: Prolate Functions, Sampling, and Applications*. Springer Science & Business Media, 2011.
[29] J. A. Hogan and J. D. Lakey. Wavelet frames generated by bandpass prolate functions. In *Int. Conf. Sampling Theory Applic. (SampTA)*, pages 120–123, 2015.
[30] A. Israel. The eigenvalue distribution of time-frequency localization operators. *arXiv:1502.04404*, 2015.
[31] S. Izu and J. D. Lakey. Time-frequency localization and sampling of multiband signals. *Acta Appl. Math.*, 107(1-3):399–435, 2009.
[32] A. Jain and S. Ranganath. Extrapolation algorithms for discrete signals with application in spectral estimation. *IEEE Trans. Acous. Speech Signal Process.*, 29(4):830–845, 1981.
[33] T. Kailath, A. Vieira, and M. Morf. Inverses of Toeplitz operators, innovations, and orthogonal polynomials. *SIAM Rev.*, 20(1):106–119, 1978.
[34] S. Karnik, J. Romberg, and M. A. Davenport. Improved bounds for the eigenvalues of prolate spheroidal wave functions and discrete prolate spheroidal sequences. *Applied and Computational Harmonic Analysis*, 55:97–128, 2021.
[35] S. Karnik, J. Romberg, and M. A. Davenport. Thomson's multitaper method revisited. *arXiv preprint arXiv:2103.11586*, 2021.
[36] S. Karnik, Z. Zhu, M. Wakin, J. Romberg, and M. Davenport. Fast computations for approximation and compression in Slepian spaces. In *Proc. IEEE. Global Conf. on Signal and Inform. Processing (GlobalSIP)*, Washington, DC, Dec. 2016.
[37] S. Karnik, Z. Zhu, M. B. Wakin, J. K. Romberg, and M. A. Davenport. The fast Slepian transform. to appear in *Appl. Comp. Harm. Anal., arXiv preprint arXiv:1611.04950*.
[38] A. Kolmogorov. Über die beste anna'aherung von funktionen einer gegebenen funktionenklasse. *Annals Math.*, 37(1):107–110, 1936.

[39] H. A. Krieger. Toeplitz operators on locally compact abelian groups. *J. Math. Mech.*, 14:439–478, 1965.
[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
[41] H. Landau. On Szegő's eingenvalue distribution theorem and non-hermitian kernels. *Journal d'Analyse Mathématique*, 28(1):335–357, 1975.
[42] H. Landau. On the density of phase-space expansions. *IEEE Trans. Inform. Theory*, 39(4):1152–1156, 1993.
[43] H. Landau and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. II. *Bell Systems Tech. J.*, 40(1):65–84, 1961.
[44] H. Landau and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. III. *Bell Systems Tech. J.*, 41(4):1295–1336, 1962.
[45] H. Landau and H. Widom. Eigenvalue distribution of time and frequency limiting. *J. Math. Anal. Appl.*, 77(2):469–481, 1980.
[46] T. J. Lim and M. Franceschcetti. A deterministic view on the capacity of bandlimited functions. In *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 691–698, 2014.
[47] T. J. Lim and M. Franceschetti. Information without rolling dice. *IEEE Trans. Inform. Theory*, 2017.
[48] I. J. Linnik. On the canonical distribution of a class of toeplitz operators on a locally compact abelian group. *Mathematics of the USSR-Izvestiya*, 10(5):1133–1142, 1976.
[49] J. Makhoul. Linear prediction: A tutorial review. *Proc. IEEE*, 63(4):561–580, 1975.
[50] R. Matthysen and D. Huybrechs. Fast algorithms for the computation of fourier extensions of arbitrary length. *SIAM J. Sci. Comput.*, 38(2):A899–A922, 2016.
[51] A. Osipov. Certain upper bounds on the eigenvalues associated with prolate spheroidal wave functions. *Appl. Comput. Harmon. Anal.*, 35(2):309–340, 2013.
[52] S. V. Parter. On the distribution of the singular values of Toeplitz matrices. *Linear Algebra Appl.*, 80:115–130, 1986.
[53] J. Pearl. On coding and filtering stationary signals by discrete Fourier transforms. *IEEE Trans. Inf. Theory*, 19(2):229–232, 1973.
[54] A. Pinkus. *N-widths in Approximation Theory*, volume 7. Springer Science & Business Media, 2012.
[55] W. Rudin. *Fourier analysis on groups*. John Wiley & Sons, 2011.
[56] D. Sakrison. An extension of the theorem of Kac, Murdock and Szegő to N dimensions. *IEEE Trans. Inf. Theory*, 15(5):608–610, 1969.
[57] H. A. A. Saleh, A. F. Molisch, T. Zemen, S. D. Blostein, and N. B. Mehta. Receive antenna selection for time-varying channels using discrete prolate spheroidal sequences. *IEEE Trans. Wire. Commun.*, 11(7):2616–2627, 2012.
[58] K. Scaman and A. Virmaux. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 3839–3848, 2018.
[59] A. Schaeffer. Inequalities of A. Markoff and S. Bernstein for polynomials and related functions. *Bulletin Amer. Math. Soc.*, 47(8):565–579, 1941.
[60] E. Sejdić, A. Can, L. Chaparro, C. Steele, and T. Chau. Compressive sampling of swallowing accelerometry signals using time-frequency dictionaries based on modulated discrete prolate spheroidal sequences. *EURASIP J. Adv. Signal Processing*, 2012(1):1–14, 2012.
[61] E. Sejdić, M. Luccini, S. Primak, K. Baddour, and T. Willink. Channel estimation using DPSS based frames. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Las Vegas, NV, Mar. 2008.
[62] S. Senay, L. F. Chaparro, and L. Durak. Reconstruction of nonuniformly sampled time-limited signals using prolate spheroidal wave functions. *Signal Process.*, 89(12):2585–2595, 2009.
[63] S. Senay, L. F. Chaparro, M. Sun, and R. J. Sclabassi. Compressive sensing and random filtering of eeg signals using slepian basis. In *16th European Signal Process. Conf.*, pages 1–5, 2008.
[64] C. E. Shannon. A mathematical theory of communication. *ACM Mobile Comput. Commun. Review*, 5(1):3–55, 2001.
[65] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. IV. *Bell Systems Tech. J.*, 43(6):3009–3058, 1964.
[66] D. Slepian. On Bandwidth. *Proc. IEEE*, 64(3):292–300, 1976.
[67] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V- The discrete case. *Bell Syst. Tech. J.*, 57(5):1371–1430, 1978.
[68] D. Slepian. Some comments on Fourier analysis, uncertainty, and modeling. *SIAM Rev.*, 25(3):379–393, 1983.
[69] D. Slepian and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. I. *Bell Systems Tech. J.*, 40(1):43–64, 1961.

[70] S. Sreepada and S. Kalyani. Channel estimation in OFDM systems with virtual subcarriers using DPSS. *IEEE Communications Letters*, 20(12):2462–2465, 2016.
[71] H. Stark and J. W. Woods. *Probability, random processes, and estimation theory for engineers*. Prentice-Hall, 1986.
[72] T. Tao. 245c, notes 2: The fourier transform, 2009.
[73] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70(9):1055–1096, 1982.
[74] V. Tikhomirov. ε-entropy and ε-capacity of sets in functional spaces. In *Selected works of AN Kolmogorov*, pages 86–170. Springer, 1993.
[75] O. Toeplitz. Zur theorie der quadratischen und bilinearen formen von unendlichvielen veränderlichen. *Mathematische Annalen*, 70(3):351–376, 1911.
[76] E. E. Tyrtyshnikov. A unifying approach to some old and new theorems on distribution and clustering. *Linear Algebra Appl.*, 232:1–43, 1996.
[77] A. W. Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge university press, 2000.
[78] C. Vonesch, F. Stauber, and M. Unser. Steerable PCA for rotation-invariant image recognition. *SIAM J. Imag. Sci.*, 8(3):1857–1873, 2015.
[79] H. Xiao, V. Rokhlin, and N. Yarvin. Prolate spheroidal wavefunctions, quadrature and interpolation. *Inverse Prob.*, 17(4):805, 2001.
[80] J. Xu. Degrees of freedom of oam-based line-of-sight radio systems. *IEEE Trans. Antenn. Prop.*, 65(4):1996–2008, 2017.
[81] X. Yang, Y. Liu, and T. Long. Robust non-homogeneity detection algorithm based on prolate spheroidal wave functions for space-time adaptive processing. *IET Radar, Sonar & Navigation*, 7(1):47–54, 2013.
[82] X. Yi. Asymptotic singular value distribution of linear convolutional layers. *arXiv preprint arXiv:2006.07117*, 2020.
[83] N. L. Zamarashkin and E. E. Tyrtyshnikov. Distribution of eigenvalues and singular values of Toeplitz matrices under weakened conditions on the generating function. *Sbornik: Mathematics*, 188(8):1191–1201, 1997.
[84] T. Zemen and C. Mecklenbräuker. Time-variant channel estimation using discrete prolate spheroidal sequences. *IEEE Trans. Signal Process.*, 53(9):3597–3607, 2005.
[85] T. Zemen, C. Mecklenbräuker, F. Kaltenberger, and B. Fleury. Minimum-energy band-limited predictor with dynamic subspace selection for time-variant flat-fading channels. *IEEE Trans. Signal Process.*, 55(9):4534–4548, 2007.
[86] Y. Zeng and Y.-C. Liang. Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Trans. Commun.*, 57(6):1784–1793, 2009.
[87] Z. Zhu, S. Karnik, M. A. Davenport, J. K. Romberg, and M. B. Wakin. The eigenvalue distribution of discrete periodic time-frequency limiting operators. *to appear in IEEE Signal Process. Letters*, 2017.
[88] Z. Zhu, S. Karnik, M. B. Wakin, M. A. Davenport, and J. Romberg. ROAST: Rapid orthogonal approximate slepian transform. *IEEE Trans. Signal Process.*, 66(22):5887–5901, 2018.
[89] Z. Zhu, S. Karnik, M. B. Wakin, M. A. Davenport, and J. K. Romberg. Fast orthogonal approximations of sampled sinusoids and bandlimited signals. In *IEEE Conf. Acous., Speech, Signal Process. (ICASSP)*, pages 4511–4515, 2017.
[90] Z. Zhu and M. Wakin. Wall clutter mitigation and target detection using discrete prolate spheroidal sequences. In *Proc. Int. Work. on Compressed Sensing Theory Appl. Radar, Sonar and Remote Sens. (CoSeRa)*, Pisa, Italy, Jun. 2015.
[91] Z. Zhu and M. B. Wakin. On the dimensionality of wall and target return subspaces in through-the-wall radar imaging. In *4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, pages 110–114, 2016.
[92] Z. Zhu and M. B. Wakin. Approximating sampled sinusoids and multiband signals using multiband modulated DPSS dictionaries. *J. Fourier Anal. Appl.*, 23(6):1263–1310, Dec 2017.
[93] Z. Zhu and M. B. Wakin. On the asymptotic equivalence of circulant and Toeplitz matrices. *IEEE Trans. Inform. Theory*, 63(5):2975–2992, 2017.

(Z. Zhu) Department of Electrical and Computer Engineering, University of Denver, Denver, CO 80208 USA

Email address: zhihui.zhu@du.edu

(M. Wakin) Department of Electrical Engineering, Colorado School of Mines, Golden, CO 80401 USA

Email address: mwakin@mines.edu