Noname manuscript No.
(will be inserted by the editor)

Time-frequency localization and sampling of multiband signals

Scott Izu · Joseph D. Lakey

Received: date / Accepted: date

Abstract This paper begins with a review of some classical work of Landau, Slepian, Pollak and Widom concerning essentially time- and bandlimited signals and ends reviewing some recent work of Candès, Romberg and Tao that places specific but probabilistic limitations on essential time- and bandlimiting for finite signals and their discrete Fourier transforms. In between we outline some conceptual bridges from the continuous, singleband setting to the finite, multiband setting and pose a number of open problems whose solutions would solidify the connections outlined. These connections involve time-frequency localization of multiband signals and sampling theory for such signals.

Keywords Sampling · multiband signal · time-frequency localization · uncertainty principle

1 Introduction

This paper reviews some connections between sampling and time-frequency localization both in the case of the real line $\mathbb{R}$ and in the finite setting, and poses several open problems. We will be concerned primarily with *multiband* signals. Fundamental issues in this context include interpolation from samples and dimensionality of the space of *essentially* time- and bandlimited signals. Practical issues include the increase in complexity that comes with dispersion of frequency support and the extent to which a finite signal spectrum can serve as a model for the spectrum of an analogue signal. We will only address these practical points superficially but they underlie the problems relating sampling and time-frequency localization that are posed here. To help focus

S. Izu
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003-8001 E-mail: scottizu@gmail.com

J. Lakey
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003-8001 E-mail: jlakey@nmsu.edu

on technical methods, we will systematically avoid several importantly related topics. Specifically, we will not discuss phase-space counterparts of the time-frequency theory or reproducing kernel Hilbert space counterparts of the sampling theory.

The paper is organized as follows. We begin reviewing the work of Landau, Slepian and Pollak [35,28,29,32] – the so-called *Bell Labs* theory – concerning "time- and bandlimiting" (see also [22,24,34,26]) in Section 2. The Paley-Wiener space PW consists of those functions in $L^2(\mathbb{R})$ whose Fourier transforms vanish outside the set $[-1/2,1/2]$. We normalize the Fourier transform so that $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-2\pi it\xi}\,dt$ for $f \in L^1 \cap L^2(\mathbb{R})$. For $f \in$ PW one can ask how much of the energy of f can be concentrated in a given time interval I. This energy concentration depends only on $|I|$. Let P denote the orthogonal projection onto PW and denote by Q_T the operator $Q_T f(x) = f(x)\mathbb{1}_{[-T/2,T/2]}(x)$ where $\mathbb{1}_S$ denotes the indicator function taking the value one on S and zero off S. The energy concentration problem amounts to that of finding the norm of the self-adjoint operator PQ_TP and was addressed in the Bell Labs theory. Subsequent work of Landau and Widom addressed eigenvalue asymptotics in the corresponding case of time- and bandlimiting to unions of intervals. Their work will be reviewed in some detail in Section 3. Landau and Widom's theorem (Theorem 3) says that, in an asymptotic sense, the dimension of the space of functions that are essentially time- and bandlimited to sets S (in time) and Σ (in frequency) is essentially the area $|S||\Sigma|$ when S and Σ are finite unions of intervals. Before outlining Landau and Widom's work, though, we will present some later work of Landau [27] showing that the number of eigenvalues of PQ_TP larger than $1/2$ is essentially the time-frequency area T.

The discussion of the Bell Labs theory will be followed briefly in Section 4 with an outline of issues pertaining to the connection between finite and analogue settings, including a discussion of estimates of Auslander and Grünbaum [2] quantifying the sorts of errors that arise when passing from analogue signals to certain finite models. After this we review in Section 5 work of Candès, Romberg and Tao (CRT) in [8,9] that addresses a problem somewhat parallel to that of the multiple interval eigenvalue estimates of Landau and Widom, but in the finite setting and with an emphasis on sparsity of time and frequency supports. The CRT result is essentially that, for large N, given subsets S and Σ now of $\mathbb{Z}_N$, if the counting measures $|S|$ and $|\Sigma|$ of S and Σ are small enough then, *with high probability*, the time-frequency localization matrix $A_{S\Sigma} = \mathbb{1}_S \mathcal{F}_N^{-1} \mathbb{1}_\Sigma \mathcal{F}_N \mathbb{1}_S$ will have norm less than $1/2$ (cf. [8]). Here, $\mathbb{1}_S$ denotes the operator of multiplication by the characteristic function of S and $\mathcal{F}_N$ denotes the N-point discrete Fourier transform (DFT). In this setting one can think of S as being fixed and Σ as being randomly generated from the collection of all subsets of $\mathbb{Z}_N$ of a fixed size. Candès et al. were motivated by a desire to quantify the sense in which signals having sparse frequency support can be recovered exactly from a small number of time samples according to an ℓ^1-*minimization* criterion. But their result (Theorem 6) also sheds light on the sense in which the "area formula" given by Theorem 3 breaks down when the frequency set is highly disconnected. We will review the CRT methods for estimating the norm of $A_{S\Sigma}$ in some detail in Section 5 and further in Appendix A, including further commentary on the relationship between the finite and continuous settings.

The sampling theorem is the cornerstone on which the connection between analogue signals and their discrete (infinite) counterparts is based, and the Nyquist rate provides an informational basis for this connection. The *classical* sampling theorem attributed to Shannon, Whittaker and many others (e.g., [16,14,43]) states that for any $f \in \mathrm{PW}$ one can write

$$f(t) = \sum_{k\in\mathbb{Z}} f(k)\,\mathrm{sinc}\,(t-k) \tag{1}$$

with convergence in $L^2(\mathbb{R})$.

While the Bell Labs theory addressed several aspects of time-frequency localization, it did not address potentially useful connections between time-frequency localization and sampling. Examples of such connections have been established only relatively recently, independently by Khare and George [21] and by Shen and Walter [41]. Specifically, both sets of authors showed that the integer samples of the eigenfunctions of PQ_TP are eigenvectors of the matrix $A_{k\ell} = \int_{-T/2}^{T/2} \mathrm{sinc}\,(t-k)\,\mathrm{sinc}\,(t-\ell)\,dt$. We extend this relationship to multiband signals in Section 6, but the extension is contingent upon having a method by which such signals can be interpolated from their samples.

A sampling formula corresponding to (18) applies to any f such that $\hat{f}$ is supported in a closed and bounded interval. The Nyquist sampling rate is the reciprocal of the length of this interval. Several approaches to effective sampling of multiband signals have been proposed in recent years, e.g., [13,40,3]. Among them, works of Venkataramani and Bresler and of Behmard and Faridani establish formulas or algorithms for interpolating multiband signals from samples and their methods will be reviewed in Section 7. Filterbank based approaches due to Herley and Wong [13] and Eldar and Oppenheim [12] will not be discussed. The common goal of these methods is to reproduce signals from samples taken at an average rate inversely proportional to the measure of the frequency support. The contributions just noted are far from an exhaustive list of the work done on this front; see, e.g., Higgins [15] for earlier work. In the subsequent discussion some basic background from spectral theory that can be found, for example, in Arveson's book [1] will be assumed.

2 Time- and bandlimiting to pairs of intervals

One form of the uncertainty principle says that there is no $f \in L^2(\mathbb{R}) \setminus \{0\}$ such that $\{t:\ f(t) \neq 0\}$ and $\{\xi:\ \hat{f}(\xi) \neq 0\}$ both have finite measure. Given sets S and Σ of finite measure, the operators $P_\Sigma f = (\hat{f}\chi_\Sigma)^\vee$ and $Q_S = f\chi_S$ are both self-adjoint projection operators. They are also both compact, having traces equal to their Lebesgue measures $|S|$ and $|\Sigma|$ respectively. But no nontrivial $f \in L^2(\mathbb{R})$ is fixed by both Q_S and P_Σ [5]. Nonetheless, one can still consider the extent to which f can be jointly localized on the pair S, Σ. A deep theorem due to Nazarov [31] provides the following bound for this localization,

$$\int_S |f|^2 + \int_\Sigma |\hat{f}|^2 \leq C e^{A|S||\Sigma|}\left(\int_{\mathbb{R}\setminus S} |f|^2 + \int_{\mathbb{R}\setminus\Sigma} |\hat{f}|^2\right)$$

with fixed constants A and C. The case of the Gaussian $\mathrm{e}^{-\pi x^2}$ shows that the estimate is asymptotically optimal.

In the remainder of this section we will review briefly several properties of the time-frequency localization operators proved by Landau, Slepian and Pollak in the case S and Σ are intervals, and raise some questions regarding extensions to more general cases.

Bandlimiting and interpolation. If Σ is compact then $\mathrm{PW}_\Sigma = P_\Sigma L^2$ is a space of entire functions of exponential type. In particular, any $f \in \mathrm{PW}_\Sigma$ can be recovered from its values on a particular set of samples, provided the sampling set is *dense enough* (see Landau [25]). Finding functions that interpolate all $f \in \mathrm{PW}_\Sigma$ from a given sampling set is a nontrivial matter, in general, but useful interpolating formulas are known in several cases.

The Paley-Wiener space and the sampling theorem. Abusing notation, for $\Omega > 0$ fixed we will denote by $\mathrm{PW}_\Omega = \mathrm{PW}_{[-\Omega/2,\Omega/2]}$ the *Paley-Wiener space* $P_\Omega L^2(\mathbb{R})$ where $P_\Omega = P_{[-\Omega/2,\Omega/2]}$. The space PW_Ω is a reproducing kernel Hilbert space with kernel $K(s,t) = \mathrm{sinc}_\Omega(t-s)$ where

$$\mathrm{sinc}_\Omega(t) = \frac{\sin(\pi\Omega t)}{\pi t} = \Omega\,\mathrm{sinc}\,(\Omega t). \tag{2}$$

The reproducing property

$$\int_{\mathbb{R}} \mathrm{sinc}_\Omega(t-v)\,\mathrm{sinc}_\Omega(v-s)\,dv = \mathrm{sinc}_\Omega(t-s). \tag{3}$$

follows directly from Parseval's identity and the fact that $(\mathbb{1}_{[-\Omega/2,\Omega/2]})^\vee(t) = \mathrm{sinc}_\Omega(t)$.

A simple calculation shows that for $f \in \mathrm{PW}_\Omega$, the n-th Fourier coefficient of the Fourier series expansion of $\widehat{f}$ on $[-\Omega/2,\Omega/2]$ is $f(-n/(2\Omega))/\Omega$. The Fourier inversion formula then yields the sampling formula:

$$f(t) = \frac{1}{\Omega}\sum_{k\in\mathbb{Z}} f\Big(\frac{k}{\Omega}\Big)\mathrm{sinc}_\Omega\Big(t-\frac{k}{\Omega}\Big). \tag{4}$$

Unfortunately, sinc_Ω is poorly localized in time so, generally, many samples are needed to approximate f. This raises the natural question: what additional conditions on $f \in \mathrm{PW}_\Omega$ allow for good time-localized sampling approximations? The answer is partly that f should also be well localized in time.

Joint localization to a time interval and a frequency interval. For $S \subset \mathbb{R}$ let Q_S denote the operator of multiplication by $\mathbb{1}_S$ and, abusing notation as before, for $T > 0$ let $Q_T = Q_{[-T/2,T/2]}$. The composition $P_\Omega Q_T$ is (compact and) self-adjoint when considered as an operator on PW_Ω. Denote by $\lambda_0 \geq \lambda_1 \cdots \geq 0$ and $\varphi_n \in \mathrm{PW}_\Omega$ respectively the eigenvalues and corresponding normalized eigenfunctions, $P_\Omega Q_T \varphi_n = \lambda_n \varphi_n$. Loosely speaking, a function is (T,Ω) time-frequency localized if it is a sum of eigenfunctions φ_n of $P_\Omega Q_T$ having eigenvalues close to one. It is important then to quantify the manner in which those eigenvalues decay.

Double orthogonality for $P_\Sigma Q_S$. Since $P_\Omega Q_T$ is self-adjoint on PW_Ω, its (normalized) eigenfunctions $\{\varphi_n\}$ form an orthonormal basis for PW_Ω. In fact, the functions $Q_T\varphi_n$ are also orthogonal. For the moment let S and Σ be any compact subsets of $\mathbb{R}$. In [35] Slepian and Pollak observed that orthogonality of eigenfunctions of $P_\Sigma Q_S$ over S really just depends on a reproducing property and symmetry of the kernel.

Theorem 1 *Suppose that ρ is even and satisfies*

$$\int \rho(t-u)\,\rho(u-s)\,du = \rho(t-s),$$

and that the eigenfunctions of $T_{\rho,S} : f \mapsto \int_S \rho(t-u)\,f(u)\,du$ are in $L^2(\mathbb{R})$. If φ_λ and φ_μ are eigenfunctions of $T_{\rho,S}$ with eigenvalues λ and μ respectively then

$$\int_{-\infty}^{\infty} \varphi_\lambda(t)\,\varphi_\mu(t)\,dt = \frac{1}{\lambda}\int_S \varphi_\lambda(t)\,\varphi_\mu(t)\,dt.$$

In particular, if $\lambda \neq \mu$ then the eigenfunctions are orthogonal over S.

Proof First we note that T_ρ could have degenerate eigenvalues, though we will ignore this case. One has

$$\begin{aligned}
\int_{-\infty}^{\infty} \varphi_\lambda(t)\,\varphi_\mu(t)\,dt &= \frac{1}{\lambda\mu}\int_{-\infty}^{\infty}\int_S \rho(t-s)\,\varphi_\lambda(s)\,ds \int_S \rho(t-u)\,\varphi_\mu(u)\,du\,dt \\
&= \frac{1}{\lambda\mu}\int_S \varphi_\lambda(s)\int_S \varphi_\mu(u)\int_{-\infty}^{\infty} \rho(t-s)\,\rho(t-u)\,dt\,ds\,du \\
&= \frac{1}{\lambda\mu}\int_S \varphi_\lambda(s)\int_S \varphi_\mu(u)\,\rho(s-u)\,du\,ds \\
&= \frac{1}{\lambda}\int_S \varphi_\lambda(s)\,\varphi_\mu(s)\,ds.
\end{aligned}$$

When $\lambda \neq \mu$ one can interchange the roles of λ and μ to conclude that

$$\int_{-\infty}^{\infty} \varphi_\lambda(t)\,\varphi_\mu(t)\,dt = \frac{1}{\lambda}\int_S \varphi_\lambda(s)\varphi_\mu(s)\,ds = \frac{1}{\mu}\int_S \varphi_\lambda(s)\varphi_\mu(s)\,ds = 0.$$

The hypotheses on ρ apply whenever $\widehat{\rho}$ is a symmetric linear combination of characteristic functions of bounded intervals and, more generally, when $\widehat{\rho} = \mathbb{1}_\Sigma$ where $\Sigma = -\Sigma$. If Σ is not symmetric the kernel ρ_Σ need not be even or Hermitian and the double orthogonality may not hold then.

Energy maximization. Under the hypotheses of Theorem 1, for $f \in L^2(\mathbb{R})$ one has

$$\|P_\Sigma Q_S f\|^2 = \int_S\int_S \rho_\Sigma(t-s)\,f(t)\,\overline{f}(s)\,dt\,ds.$$

This quantity is maximized over all f by the restriction of φ_0 to S where φ_0 is the eigenfunction of $P_\Sigma Q_S$ with largest eigenvalue λ_0.

On the other hand, one can ask which $g \in \mathrm{PW}_\Sigma$ loses the least fractional energy when first timelimited to S then bandlimited to Σ. The answer again is φ_0; now the remaining fractional energy is λ_0^2, at least under the added hypothesis that λ_0 is nondegenerate. In this case the result follows from the fact that the eigenfunctions form a complete orthonormal basis for $L^2(\mathbb{R})$.

Self-reciprocal property. We return to the special case in which S and Σ are intervals. If φ is an eigenfunction of $P_\Omega Q_T$ with eigenvalue λ then its Fourier transform $\widehat{\varphi}$ satisfies

$$\widehat{\varphi}\Big(\frac{\Omega}{T}\xi\Big) = i^p \sqrt{\frac{T}{\Omega}}\,\lambda\, Q_T\varphi(\xi). \tag{5}$$

Here p is an integer that depends only on the magnitude ordering of the eigenvalues. We refer to [18] for the simple calculation that verifies (5). One concludes that the Fourier transform of an eigenfunction is essentially just the cutoff of a dilate of the eigenfunction. This property obviously does not extend to general sets S and Σ, though it will extend to cases in which $S = -S$ is a dilate of Σ.

Actually, since $\widehat{\alpha\varphi(\alpha\cdot)}(\xi) = \widehat{\varphi}(\xi/\alpha)$, the operator $P_\Omega Q_T$ is simply a dilation of $P_1 Q_{T\Omega}$. Thus if we set $c = T\Omega$ it suffices to study the operators PQ_c where $P = P_1$. We will write $\lambda_j = \lambda_j(c)$ when dependence of the eigenvalue on c is to be emphasized. We call c the *time-frequency area*.

Prolate spheroidal wave functions. Slepian and Pollak [35] observed the "lucky accident" that eigenfunctions of $Q_c P$ also happen to be solutions of

$$(1-t^2)\frac{d^2u}{dt^2} - 2t\frac{du}{dt} + (\gamma - c^2t^2)u = 0$$

for corresponding values of γ. The latter are the so-called *prolate spheroidal wave functions* (PSWFs) that are well-studied through their physical applications [37,30] Of course, this fact will not generalize at all to arbitrary sets S and Σ. In general, eigenfunctions of $P_\Sigma Q_S$ have to be estimated numerically.

Completeness. One consequence of the identification of the eigenfunctions of PQ_c as PSWFs is their completeness over $[-1/2, 1/2]$. However, this completeness is also a consequence of (5) as was observed by Xiao, Yarvin and Rokhlin [42]. Writing the sinc kernel in terms of the eigenfunctions φ_j and applying (5) with suitable normalizing factors α_j gives

$$\operatorname{sinc}_c(\xi-\eta) = \sum_{j=0}^{\infty} \varphi_{j,c}(\xi)\,\varphi_{j,c}(\eta) = \Big(\int_{-c/2}^{c/2} \mathrm{e}^{-2\pi i x\xi} \sum_{j=0}^{\infty} \frac{1}{\alpha_j}\varphi_{j,c}(\xi)\,dx\Big)\,\varphi_{j,c}(c\eta).$$

Applying Fourier inversion one concludes that

$$c\sum_{n=0}^{\infty} \frac{1}{\alpha_n}\varphi_n(x)\varphi_n\big(c\eta\big) = \mathrm{e}^{2\pi i\eta x}, \quad (|x| < c/2).$$

Fourier uniqueness now implies that the functions φ_n are complete in $L^2[-c/2, c/2]$ since, if $f \in L^2[-c/2, c/2]$ is orthogonal to each of the φ_n then it is orthogonal (on the interval) to each $\mathrm{e}^{2\pi i x\xi}$ for each ξ, hence $f = 0$ on $[-c/2, c/2]$, proving completeness.

If Σ is a finite union of intervals, completeness of the eigenfunctions of $P_\Sigma Q_T$ in $L^2[-T/2, T/2]$ can be established by reasoning along the lines of Proposition 2 below.

3 Eigenvalues for time-frequency localization

3.1 The number of eigenvalues of $P_\Omega Q_T$ larger than 1/2

In their works [29, 34, 24], Landau, Slepian, Pollak and Widom proved a number of statements of the "ΩT" stating, in essence, that the dimension of the space of essentially T-timelimited and Ω-bandlimited signals is essentially the time-frequency area ΩT. One version–Theorem 3 below–says that $P_\Omega Q_T$ has about ΩT eigenvalues close to one, and that the eigenvalues plunge rapidly from $\lambda \approx 1$ to $\lambda \approx 0$ over a transition band of width around $\log \Omega T$. In particular, that result implies that for large ΩT, the number of eigenvalues of $P_\Omega Q_T$ greater than 1/2 is $\Omega T + o(\log \Omega T)$. In [27], (cf. [22]) Landau provides the following simple and more precise version of this last fact.

Theorem 2 *Let* $Q_c f = f \mathbb{1}_{[-c/2,c/2]}$ *and* $Pf = (\widehat{f}\,\mathbb{1}_{[-1/2,1/2]})^\vee$. *The eigenvalues of* PQ_cP *satisfy*

$$\lambda_{\lfloor c \rfloor -1} \geq 1/2 \geq \lambda_{\lceil c \rceil}.$$

In the theorem, $\lfloor x \rfloor$ and $\lceil x \rceil$ refer the greatest integer less than or equal to x and least integer greater than or equal to x respectively. The Weyl-Courant minimax characterization of the singular values $\lambda_0 \geq \lambda_1 \geq \dots$ of $P_\Sigma Q_S P_\Sigma$ can be stated as

$$\lambda_k = \begin{cases} \min_{\mathcal{S}_k} & \max\{\|Q_S f\|^2 : f \in \mathrm{PW}_\Sigma, \|f\| = 1,\ f \perp \mathcal{S}_k\} \\ \max_{\mathcal{S}_{k+1}} & \min\{\|Q_S f\|^2 : f \in \mathrm{PW}_\Sigma, \|f\| = 1,\ f \in \mathcal{S}_{k+1}\} \end{cases}.$$

Here, $\mathcal{S}_k$ ranges over all k-dimensional subspaces including, notably, the subspace spanned by the first k eigenfunctions of $P_\Sigma Q_S P_\Sigma$. In [22] Landau identifies a convolver h such that if $f \in \mathrm{PW}$ and $f * h(m)$ vanishes at a given set of $k = \lceil c \rceil$ integers then $\|Q_c f\|^2 \leq 0.6$. The bound was improved to $\|Q_c f\|^2 \leq 1/2$ in [27]. One can verify numerically the finite version of Theorem 2 for the so-called *discrete prolate spheroidal sequences* of order N and bandwidth W, defined as the eigenvectors of the matrix $M_{N,W}(n,m) = \mathrm{sinc}_{\,2W}(n-m)$, $n, m = 0, 1, \dots, N-1$. In this case the *discrete normalized area* amounts to $2W$ and one observes $\lambda_{2W-1} > 1/2 > \lambda_{2W}$ when $W \in \{1/2, 1, \dots, (N-1)/2\}$.

Proof Assume for the moment that $c \notin \mathbb{N}$. If h is supported in $[-1/2, 1/2]$ then

$$f * h(x) = \int_{x-1/2}^{x+1/2} f(t) h(x-t)\, dt$$

and

$$|f * h(k)|^2 \leq \|h\|^2 \int_{k-1/2}^{k+1/2} |f(t)|^2\, dt.$$

Choose I to be any closed interval of length c chosen so that the interval I_+ obtained by extending I by 1/2 on each end contains as few integers as possible, namely $\lceil c \rceil$ of them. Let $\mathcal{S}_+$ denote the closed span of the functions $h(k - \cdot)$, $k \in I_+$. If $f \perp \mathcal{S}_+$ then $f * h(k) = 0$ when $k \in I_+$ so

$$\begin{aligned} \sum |f * h(k)|^2 &= \sum_{k \notin I_+} |f * h(k)|^2 \leq \|h\|^2 \int_{t \notin I} |f(t)|^2 \\ &= \|h\|^2 \big(\|f\|^2 - \|Q_c(f)\|^2\big). \end{aligned}$$

On the other hand, since $f * h \in \mathrm{PW}$ when f is,

$$\sum |f * h(k)|^2 = \int_{-1/2}^{1/2} |\hat{f}(\xi)\hat{h}(\xi)|^2 \, d\xi$$
$$\geq \alpha^2 \int_{-1/2}^{1/2} |\hat{f}(\xi)|^2 \, d\xi = \alpha^2 \|f\|^2$$

provided that $|\widehat{h}(\xi)| \geq \alpha$ on $[-1/2, 1/2]$. In this case, if $f \perp \mathcal{S}_+$ then

$$\|Q_c(f)\|^2 \leq \|f\|^2 \Big(1 - \frac{\alpha^2}{\|h\|^2}\Big).$$

The dimension of $\mathcal{S}_+$ is the number of integers in I_+ which is $\lceil c \rceil$. Thus, if $f \perp \mathcal{S}_+$ then $\|Q_c f\|^2/\|f\|^2 \leq (1-\alpha^2/\|h\|^2) \leq 1/2$ provided that one can find h having $\|h\| = 1$ such that $|\widehat{h}(\xi)| \geq 1/\sqrt{2}$ on $[-1/2, 1/2]$. Recall that h is supported in $[-1/2, 1/2]$. Taking $h = \sqrt{2}\cos \pi t$ on $[-1/2, 1/2]$ and $h = 0$ elsewhere gives $\|h\|^2 = 2\int_{-1/2}^{1/2} \cos^2 \pi t \, dt = 1$ whereas $\widehat{h}(\xi) = \frac{1}{\sqrt{2}}(\mathrm{sinc}\,(\xi+1/2) + \mathrm{sinc}\,(\xi-1/2))$. Then $\widehat{h}(-1/2) = \widehat{h}(1/2) = 1/\sqrt{2}$ and concavity of $\widehat{h}$ on $[-1/2, 1/2]$ implies that $h(\xi) \geq 1/\sqrt{2}$ on $[-1/2, 1/2]$. One concludes from the minimax criterion that $\lambda_{\lceil c \rceil} \leq 1/2$. In the limiting case $c \in \mathbb{N}$ the same estimate follows from the fact that for each $k = 0, 1, \ldots,$ $\lambda_k(c)$ varies continuously with c.

For the other direction we will assume that $c > 1$. Let I be an interval of length c chosen so that the interval I_- obtained by shortening I by $1/2$ at each end still contains $\lfloor c \rfloor$ integers. With h as above, let $q \in \mathrm{PW}$ be such that $\widehat{q}(\xi) = 1/\overline{\widehat{h}}(\xi)$ on $[-1/2, 1/2]$. Let $\mathcal{S}_-$ be the span of $q(\cdot - k)$ where k ranges over the integers in I_-. If $f \in \mathcal{S}_-$ then we can write

$$\widehat{f}(\xi) = \sum_{k \in I_-} b_k \mathrm{e}^{-2\pi i k \xi} / \overline{\widehat{h}}(\xi)$$

so that

$$\widehat{f}(\xi)\overline{\widehat{h}}(\xi) = \sum_{k \in I_-} b_k \mathrm{e}^{-2\pi i k \xi}, \quad |\xi| \leq 1/2.$$

Then

$$\sum_{k \in I_-} |b_k|^2 = \int_{-1/2}^{1/2} |\widehat{f}(\xi)\hat{h}(\xi)|^2 \, d\xi \geq \frac{1}{2}\int_{-1/2}^{1/2} |\widehat{f}|^2 = \frac{1}{2}\|f\|^2.$$

On the other hand, as before we get

$$b_k = \int_{-1/2}^{1/2} \widehat{f}(\xi)\overline{\widehat{h}}(\xi)\mathrm{e}^{2\pi i k \xi} \, d\xi = \int f(t)\overline{h(t-k)} \, dt.$$

Since h is supported in $[-1/2, 1/2]$ it follows that

$$\sum_{k \in I_-} |b_k|^2 \leq \|h\|^2 \sum_{k \in I_-} \int_{k-1/2}^{k+1/2} |f(t)|^2 = \|Q_c f\|^2$$

since $\|h\| = 1$. Altogether this shows that if $f \in \mathcal{S}_-$ then

$$\frac{1}{2}\|f\|^2 \leq \|Q_c(f)\|^2.$$

The dimension of $\mathcal{S}_-$ is $\lfloor c \rfloor$. Thus the minimax criterion tells us that $\lambda_{\lfloor c \rfloor - 1} \geq 1/2$.

Extension to multiple intervals. Landau's technique can be extended to the case in which S is a finite union of intervals (but still $\Sigma = [-1/2, 1/2]$). In [22] Landau showed that if S is a union of m intervals then

$$\begin{aligned} |S| - 2m \le \#\{k : (k-1/2, k+1/2) \subset S\} = \dim \mathcal{S}_- \\ \le \#\{k : (k-1/2, k+1/2) \cap S \neq \emptyset\} = \dim \mathcal{S}_+ \le |S| + 2m \end{aligned}$$

by letting S_+ be the union of the extension by 1/2 on each side of the intervals comprising S, letting $\mathcal{S}_+ = \operatorname{span}\{h(\cdot - k) : k \in \mathcal{S}_+\}$, and defining $\mathcal{S}_-$ in a corresponding way. Since one is unable to shift the intervals independently, this method concedes two eigenvalues bigger than 1/2 per interval. Thus $\lambda_{\lfloor c\rfloor - 2m} \ge 1/2 \ge \lambda_{\lceil c\rceil + 2m}$. However, a modest but significant gain can be achieved when the intervals all have length at least one by carefully examining the structure of Landau's proof of Theorem 2. In Izu's dissertation [19] one can find a proof of an equivalent version of the following.

Proposition 1 *Let $\Sigma = [-1/2, 1/2]$ and let S be a finite union of m pairwise disjoint intervals of total length c. Denote by $\nu = \max_\alpha \#\{k \in \mathbb{Z} : (k-1/2, k+1/2) \subset S + \alpha\}$ and $\mu = \min_\beta \#\{\ell \in \mathbb{Z} : (\ell - 1/2, \ell + 1/2) \cap S + \beta \neq \emptyset\}$. Then the eigenvalues λ_k of $Q_S P$ satisfy*

$$\lambda_{\nu-1} \ge 1/2 \ge \lambda_\mu. \tag{6}$$

In particular for $c \ge 1$, $\lfloor c\rfloor - 2m + 2 \le \nu \le \mu \le \lceil c\rceil + 2m - 2$ so that

$$\lambda_{\lfloor c\rfloor - 2m+1} \ge 1/2 \ge \lambda_{\lceil c\rceil + 2m-2}. \tag{7}$$

The case in which $m = 1$ provides the bound $\lambda_{\lfloor c\rfloor - 1} \ge 1/2 \ge \lambda_{\lceil c\rceil}$ which is just Landau's estimate, Theorem 2. The upper and lower bounds of (7) are not close to the sharper (6) in a lot of cases as the following heuristic argument indicates. Suppose that $I_1, \dots, I_m$ are the m pairwise disjoint intervals comprising S and that each has length at least one. For each j then $\lambda_{\lfloor |I_j|\rfloor - 1}(PQ_{I_j}) \ge 1/2$. Since the I_j are pairwise disjoint, from the lower bound it is safe to assume the corresponding eigenfunctions from separate intervals are linearly independent. Let $M = \sum \lfloor |I_j|\rfloor$. Then we have a subspace of $L^2(\mathbb{R})$ of dimension at least $M - m$ such that $\|Q_S P\varphi\| \ge \frac{1}{2}\|\varphi\|$ whenever φ is in this space. Thus $\lambda_{M-m-1}(PQ_S) \ge 1/2$. This inequality improves $\lambda_{\lfloor c\rfloor - 2m+1} \ge 1/2$, at least when $\sum |I_j| - \lfloor |I_j|\rfloor \ll m$, that is, $c \ll M + m$. A particularly interesting case (see [19]) is that in which each of the intervals comprising S has the form $[k - 1/2, k + 1/2)$. Then $\nu = \mu = c$ and one recovers the bounds $\lambda_{\lfloor c\rfloor - 1} \ge 1/2 \ge \lambda_{\lceil c\rceil}$ even though S can be disconnected. Of course, there may not be any eigenvalues *close to one* in this case. A pertinent question at this stage is: can one do any better than $\lambda_{M-m-1} \ge 1/2$, at least for a large class of cases. Further remarks along these lines can be found after Problem 1.

3.2 Landau and Widom's results in the case of multiple intervals

When $c = \Omega T$ the operator $P_\Omega Q_T$ corresponding to single time and frequency intervals has an eigenvalue $\lambda_{\lfloor c\rfloor} \approx 1/2$. Suppose that $T = 1$ and Σ is a finite pairwise disjoint union of c frequency intervals $I_1, \dots, I_c$ of length one, say. Then $P_\Sigma Q$ will be a complicated operator but, in view of previous observations, it will have on the order of c eigenvalues of magnitude *at least* 1/2. Consider now the limiting case in which the frequency intervals become *separated at infinity.* Any function ψ_j that is concentrated in frequency on I_j will be almost orthogonal over $[-T/2, T/2]$, in the separation limit, to any function ψ_k that is frequency concentrated on I_k when $j \neq k$. To see this, write $\psi_j(t) = \mathrm{e}^{2\pi i m_j t}\varphi_j(t)$ where m_j is the midpoint of I_j and $\widehat{\varphi_j}$ is essentially concentrated on $[-1/2, 1/2]$. Then

$$\int_{-1/2}^{1/2} \mathrm{e}^{2\pi i(m_j - m_k)t}\varphi_j(t)\overline{\varphi_k}(t)\, dt = \widehat{\varphi_1} * \widehat{\overline{\varphi_2}} * \mathrm{sinc}\,(m_1 - m_2) = O(1/|m_1 - m_2|)$$

as $|m_1 - m_2| \to \infty$. This almost orthogonality prevents eigenvalues from the separate interval operators $P_I Q$ from coalescing into significantly larger eigenvalues of $P_\Sigma Q$. Consequently $P_\Sigma Q$ will have on the order of c eigenvalues of size *approximately equal to* $1/2$ in the separation limit. Incidentally, similar reasoning shows that $P_\Sigma Q$ cannot have any eigenvalues larger than $1/2$ when Σ is a union of a large number of short, mutually distant intervals, even if $c = |\Sigma| > 1$.

Landau and Widom [24] introduced a method to estimate the distribution of eigenvalues for $P_\Sigma Q_S$ asymptotically in the sense of rescaling one of the sets S or Σ when both sets are finite unions of intervals. Their result, Theorem 3, confirmed a conjecture of D. Slepian [36] asserting that the number of eigenvalues of $P_\Omega Q_T$ is approximately $c = \Omega T$ when c is large.

Theorem 3 *Suppose that S and Σ are finite pairwise disjoint unions of N_S and N_Σ intervals respectively. Set $A_c = A_c(S, \Sigma) = P_{c\Sigma} Q_S P_{c\Sigma}$. Then the number $N(A_c, \alpha)$ of eigenvalues of A_c larger than α satisfies*

$$N(A_c, \alpha) = c|S||\Sigma| + \frac{N_S N_\Sigma}{\pi^2} \log\Big(\frac{1-\alpha}{\alpha}\Big) \log c + o(\log c) \qquad c \to \infty. \tag{8}$$

Although the factor $N_S N_\Sigma$ disappears when $\alpha = 1/2$, it appears prominently for other $\alpha \in (0, 1)$. The estimate (8) boils down to estimating the polynomial moments of the spectral measure $d_t[-N(A_c, t)]$ such that

$$N(A_c, \alpha) = \int_\alpha^1 d_t[-N(A_c, t)].$$

The result then follows from approximating $\mathbb{1}_{[\alpha,1]}$ by polynomials. The key step is to show that the moments of $d_t[-N(A_c, t)]$ are close to those of a familiar fixed kernel. The principal tool in this reduction is a variant of Szegő's limit theorem [20] also due to Landau [23]. Landau's limit theorem states in a precise sense that the eigenvalues of $A_r f = \int_{|y|<r} p(x - y)\, f(y)\, dy$ approach the Fourier transform of the kernel as $r \to \infty$. It is, perhaps, the principal reason why the case $N_S = N_\Sigma = 1$ of Theorem 3 is necessarily asymptotic. The case of finitely many time and frequency intervals involves a reduction to the single interval case which also requires asymptotic separation of the intervals.

Single interval case. Here is a brief outline of Landau and Widom's approach in the case of a single interval. The first step is a simple rescaling to the case $A_c = P_{[0,c]} Q_{[0,1]} P_{[0,c]}$. In this case one wishes to establish

$$N(A_c, \alpha) = c + \frac{1}{\pi^2} \log \frac{1-\alpha}{\alpha}\, \log c + o(\log c).$$

The polynomial moments of $d_t[-N(A_c, t)]$ are the traces of the powers of A_c. These powers are not readily expressed in terms of a kernel having a familiar Fourier transform but it turns out that those of the operator $A_c(I-A_c)$ are. As such, Landau and Widom took estimates of the powers of $A_c(I-A_c)$ as a starting point. They were able then to use symmetry properties of the corresponding spectral moments to recover corresponding estimates for A_c as we outline now.

One uses idempotency and exclusion (e.g., $P_{(-\infty,0)}+P_{(1,\infty)}=I-P_{[0,1]}$) to express the Fourier transform of $[A_c(I-A_c)]^n$ as a sum of n-th powers of four operators each unitarily equivalent to the operator

$$K_c=Q_{[1,c]}P_{[0,\infty)}Q_{(-\infty,0]}P_{[0,\infty)}Q_{[1,c]},$$

plus a remainder term that remains uniformly bounded independent of c (expressed as "$O(1)$"). For the moment estimates one expresses $A_c[A_c(I-A_c)]^n$ also as the sum of four operators each unitarily equivalent to

$$Q_{[0,c]}P_{[0,1]}K_c^n+O(1).$$

In this way one obtains

$$\begin{aligned}\operatorname{tr}[A_c(I-A_c)]^n&=4\operatorname{tr}K_c^n+O(1)\\ \operatorname{tr}A_c[A_c(I-A_c)]^n&=4\operatorname{tr}Q_{[0,c]}P_{[0,\infty)}\,K_c^n+O(1)=2\operatorname{tr}K_c^n+O(1).\end{aligned}\tag{9}$$

Trace kernel estimates. The operator K_c has kernel $k_c(x,y)=\frac{1}{4\pi^2}\int_0^\infty\frac{ds}{(s+x)(s+y)}$, defined for $1\le x\le y\le c$. This can be seen using the fact that $P_{[0,\infty)}=\frac{1}{2}(I+iH)$ where H is the Hilbert transform with kernel p.v.$\frac{1}{\pi x}$ so that $Q_{(-\infty,0]}P_{[0,\infty)}Q_{[1,c]}=\frac{i}{2}Q_{(-\infty,0]}HQ_{[1,c]}$. The change of variables $x=\mathrm{e}^{2u}$, $y=\mathrm{e}^{2v}$ and $s=\mathrm{e}^{2w}$ transforms K_c unitarily into the operator on $L^2[0,(\log c)/2]$ given by integration against $\kappa_c(s)=\frac{1}{4\pi^2}\int_{-\infty}^\infty\operatorname{sech}r\operatorname{sech}(s-r)\,dr$.

Szegő's theorem [23] can be phrased as the statement that the number of eigenvalues of K larger than α is equal to the measure of the set on which the Fourier transform of the kernel of K is larger than α. In the present context this translates into

$$N(K_c,\alpha)=\frac{\log c}{2}\big(|\{\xi:|\widehat{\kappa_c}(\xi)|>\alpha\}|+o(1)\big).$$

The $\log c$ term comes from the time support of κ_c. Since $\operatorname{sech}^\wedge(\xi)=\pi\operatorname{sech}(\pi^2\xi)$,

$$N(K_c,\alpha)=\frac{\log c}{2}|\{\xi:\operatorname{sech}^2(\pi^2\xi)>4\alpha\}|+o(\log c)=\frac{\log c}{\pi^2}\operatorname{sech}^{-1}(\sqrt{4\alpha})+o(\log c).$$

This tells us that

$$\operatorname{tr}K_c^n=\int_0^{1/4}x^n d_x[-N(K_c,x)]=\frac{\log c}{2\pi^2}\int_0^{1/4}x^n\frac{dx}{x\sqrt{1-4x}}+o(\log c).$$

Letting $x=t(1-t)$ with $0<t\le 1/2$ and using symmetry then gives

$$\operatorname{tr}K_c^n=\frac{\log c}{2\pi^2}\int_0^{1/2}(t(1-t))^n\frac{dt}{t(1-t)}+o(\log c)=\frac{\log c}{4\pi^2}\int_0^1(t(1-t))^n\frac{dt}{t(1-t)}+o(\log c).$$

Applying similar reasoning to the operators corresponding to $A_c[A_c(I-A_c)]^n$ yields a corresponding moment involving $t^{n+1}(1-t)^n$.

The spectral measure. To obtain estimates on $N(A_c, \alpha)$ now one uses its definition and (9) to write

$$\begin{aligned} \mathrm{tr}[A_c(I-A_c)]^n &= \frac{\log c}{\pi^2}\int_0^1 (t(1-t))^n \frac{dt}{t(1-t)} + o(\log c) \\ \mathrm{tr}A_c[A_c(I-A_c)]^n &= \frac{2\log c}{\pi^2}\int_0^1 t(t(1-t))^n \frac{dt}{t(1-t)} + o(\log c) \end{aligned}$$

In particular, each polynomial P that vanishes at 0 and 1 satisfies

$$\int_0^1 P(t) d_t[-N(A_c,t)] = \frac{\log c}{\pi^2}\int_0^1 P(t)\frac{dt}{t(1-t)} + o(\log c)$$

and, for any P vanishing at $t=0$,

$$\int_0^1 P(t)\, d_t[-N(A_c,t)] = P(1)\mathrm{tr}(A_c) + \frac{\log c}{\pi^2}\int_0^1 (P(t)-tP(1))\frac{dt}{t(1-t)} + o(\log c).$$

A polynomial approximation argument yields

$$N(A_c,\alpha) = \int_\alpha^1 d_t[-N(A_c,t)] = c + \frac{\log c}{\pi^2}\log\frac{1-\alpha}{\alpha} + o(\log c).$$

Extension to multiple intervals. To simplify the reduction to the single interval case we assume that Σ is a union of N_Σ bounded intervals but that S is a single time interval T. The general case (S is also a union of intervals) poses no additional obstacles. Let $\Sigma = \cup_j \Gamma_j$ with $\Gamma_j = [\alpha_j, \beta_j]$. Let $\Delta_0 = (-\infty, \alpha_0)$, $\Delta_1 = (\beta_1, \alpha_2)$, $\ldots$, $\Delta_{N_\Sigma} = (\beta_{N_\Sigma}, \infty)$ be the complementary intervals. As before, consider

$$A_c(I-A_c) = P_\Sigma Q_{cT} P_{\mathbb{R}\setminus\Sigma} Q_{cT} P_\Sigma = \sum P_{\Gamma_i} Q_{cT} P_{\Delta_j} Q_{cT} P_{\Gamma_k}$$

The individual terms in this sum are $O(1)$ (uniformly bounded independent of c) unless $i=k$ and Γ_k is adjacent to Δ_j (see the Lemma in [24]). Consequently one can express

$$A_c(I-A_c) = \sum P_{\Gamma_j} Q_{cT}(P_{(-\infty,\alpha_j)} + P_{(\beta_j,\infty)}) Q_{cT} P_{\Gamma_j} + O(1).$$

Since the result is asymptotic we can rescale so that $T=[0,1]$ and all the frequency intervals have length greater than one. Then as in the single interval case one finds that $(A_c(I-A_c))^n$ can be written as

$$\begin{aligned} [A_c(I-A_c)]^n = \sum_j &\big(P_{c[\alpha_j+1,\beta_j]} Q_{[0,\infty)} P_{c(-\infty,\alpha_j]} Q_{[0,\infty)} P_{c[\alpha_j+1,\beta_j]}\big)^n \\ &+ 3 \text{ similar sums} + O(1) \end{aligned}$$

where $cI = \{cx : x \in I\}$. As in the one interval case, to each fixed j there correspond four operators each unitarily equivalent to

$$K_{c(\beta_j-\alpha_j)} = Q_{[1,c(\beta_j-\alpha_j))} P_{(0,\infty)} Q_{(-\infty,0)} P_{(0,\infty)} Q_{[1,c(\beta_j-\alpha_j))}$$

so that

$$\operatorname{tr}[A_c(I-A_c)]^n = 4\sum_j \operatorname{tr} K^n_{c(\beta_j-\alpha_j)} + O(1) = 4\sum_j \operatorname{tr} K^n_c + O(1)$$

and one finds that

$$\operatorname{tr}[A_c(I-A_c)]^n = 4N_\Sigma \operatorname{tr} K^n_c + O(1).$$

Also as before, a similar reduction applies to $A_c[A_c(I-A_c)]^n$. The remainder of the argument follows the single interval case.

The "$O(1)$" terms can be substantial in the non asymptotic regime for two reasons. First, Szegő's theorem gives only a rough approximation. Second, the lengths and positions of the time and frequency intervals can make the "$O(1)$" term dominant in the eigenvalue estimates. These concerns lead us to raise a couple of very basic questions.

Problem 1 (i) Prove that among all compact time and frequency supports S, Σ with $|S||\Sigma| = c$, the $P_\Sigma Q_S P_\Sigma$ operator with the largest norm is (up to dilations and time-frequency shifts) $P_{[0,c]}Q_{[0,1]}P_{[0,c]}$ and (ii) quantify precisely the "$O(1)$" term in (8).

The intuition that joint localization for a given area should be optimized when S and T are intervals is further confirmed by an inequality for rearrangements due to Donoho and Stark [11]. It states that if $\Omega T < 0.8$ and if f is supported on a set of Lebesgue measure T then its symmetric, decreasing rearrangement satisfies

$$\int_{-\Omega/2}^{\Omega/2} |\widehat{f}(\xi)|^2\, d\xi \le \int_{-\Omega/2}^{\Omega/2} |(|f|^\star)^\wedge(\xi)|^2\, d\xi.$$

Since $|f|^\star$ is concentrated in $[-|S|/2, |S|/2]$ when f is supported in a set of measure $|S|$, this indicates an affirmative answer to the problem when Σ is an interval and $|S||\Sigma|$ is not too big. A more general version of Problem 1 would ask whether other eigenvalues beyond λ_0 also decrease with dispersion (i.e. disconnectedness) of S or Σ for fixed c.

As indicated earlier, in certain situations a more precise quantification of the eigenvalue distribution and possibly even an improvement of the Landau estimate [22] $\lambda_{\lfloor c-2N_S+1\rfloor} \ge 1/2 \ge \lambda_{\lceil c+2N_S-2\rceil}$ should be obtainable when Σ is an interval. A natural starting point would be the case in which all frequency intervals have the same length. Recall that $\lambda_{\lfloor c-1\rfloor} \ge 1/2 \ge \lambda_{\lceil c\rceil}$ when $S = \alpha + \cup[k, k+1)$ and $\Sigma = \beta + [0,1)$ but the situation is more complicated when the intervals comprising S do not have unit length.

Problem 2 (i) Quantify $N(A_c, \alpha)$ when $S = [0,1]$ and $\Sigma = \cup_{k=1}^M [k\beta, k\beta + L]$ with β and L fixed. (ii) Describe precise, nonasymptotic conditions on finite unions Σ for which the Landau and Widom results are essentially sharp. Specifically, give precise estimates on the $o(\log c)$ term in such cases.

3.3 Time-frequency localization for unions

Sums of time-frequency localization operators. One cannot refer to a Sturm-Liouville system for computation of the eigenvectors of $P_\Sigma Q_S$ except in the special case in which Σ and S are intervals. So it is necessary to find effective computational algorithms for estimating the eigenfunctions, for example, when Σ is a finite union of intervals. Suppose that one knows the eigenvalues and eigenfunctions for $P_{\Sigma_1}Q_S$ and $P_{\Sigma_2}Q_S$ for the same time support S and compact, pairwise disjoint frequency support sets Σ_1 and Σ_2. The following proposition [17] describes a *discrete* method for obtaining from these the eigenfunctions of $P_{\Sigma_1\cup\Sigma_2}Q_S$.

Proposition 2 *Suppose that Σ_1 and Σ_2 are disjoint, compact sets and that the eigenvectors $\{\varphi_n^{\Sigma_i}\}$ of $P_{\Sigma_i}Q_SP_{\Sigma_i}$ have corresponding nondegenerate eigenvalues listed in decreasing order as $\lambda_n^{\Sigma_i}$, $i=1,2$. Let Λ_{Σ_i} denote the diagonal matrix with n-th diagonal entry $\lambda_n^{\Sigma_i}$ and let Γ be the matrix with entries $\gamma_{nm}=\langle Q_S\varphi_n^{\Sigma_1},\, \varphi_m^{\Sigma_2}\rangle$. Then any eigenvector–eigenvalue pair ψ and λ for $P_{\Sigma_1\cup\Sigma_2}Q_S$ can be expressed as $\psi=\sum_{n=0}^{\infty}(\alpha_n\varphi_n^{\Sigma_1}+\beta_n\varphi_n^{\Sigma_2})$ where the vectors $\boldsymbol{\alpha}=\{\alpha_n\}$ and $\boldsymbol{\beta}=\{\beta_n\}$ together form a discrete eigenvector for the block matrix eigenvalue problem*

$$\lambda\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{pmatrix}=\begin{pmatrix}\Lambda_{\Sigma_1} & \overline{\Gamma}\\ \Gamma^T & \Lambda_{\Sigma_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{pmatrix}.$$

In other words, finding the eigenvectors (and eigenvalues) of $P_{\Sigma_1\cup\Sigma_2}\,Q_T$ boils down to determining the discrete eigenvectors (and eigenvalues) of the matrix in the theorem. The matrix Γ is also simpler when $\Sigma_1=I$ and $\Sigma_2=J$ are two intervals of the same length Ω. Then any function in $\mathrm{PW}(J)$ has the form $E\varphi$ for some $\varphi\in\mathrm{PW}(I)$, where E denotes the operator of multiplication by $\mathrm{e}^{2\pi i(\xi_J-\xi_I)t}$ with ξ_I the center of I. If we also take $S=[-T/2,T/2]$ then $\gamma_{nm}=\int_{-T/2}^{T/2}e^{2\pi i(\xi_I+\xi_J)t}\varphi_n(t)\varphi_m(t)\,dt$, where φ_n is the n-th PSWF, that is, the n-th eigenfunction of $P_{[-\Omega/2,\Omega/2]}Q_T$. Then the ΩT theorem suggests that the significant eigenvalues and corresponding eigenvectors of $P_{I\cup J}Q_T$ can be estimated effectively upon truncating the blocks of the matrix in the theorem to blocks of order $\Omega T\times\Omega T$.

Problem 3 Denote by M_{S,Σ_1,Σ_2} the matrix in the theorem and denote by $M_{S,\Sigma_1,\Sigma_2,N}$ the $2N\times 2N$ approximation obtained by truncating each block to size $N\times N$. Obtain accurate estimates for the norm of their difference when S, Σ_1 and Σ_2 are intervals.

The correlations γ_{nm} have to be computed either by quadrature or by more sophisticated methods, e.g., [6, 42] and dependence of such estimates on correlation errors should also be considered.

4 Finite models for continuous signals

4.1 Auslander and Grünbaum's error estimates

Auslander and Grünbaum [2] considered the problem of approximating a function in $L^2(\mathbb{R})$ by means of the DFT of some finite sampled signal. Since Dirac samples are not defined on L^2, one needs a sampling method that involves local averages both in time and in frequency. Thus one sets

$$f_\varphi(s)=f*\varphi(s)=\int f(s-t)\,\varphi(t)\,dt;\qquad \widehat{f_\psi}(\xi)=\hat f*\psi(\xi)$$

Now fix $T>0$, $\Omega>0$ and set $s_k=Tk/N$ and $\omega_k=\Omega k/N$ where $k=-N,\dots,N$. Denote by g the (centered) DFT of the samples of f_φ, thus

$$g(k)=\frac{1}{\sqrt{2N+1}}\sum_{j=-N}^{N} f_\varphi\Big(\frac{Tj}{N}\Big)\mathrm{e}^{\frac{2\pi i}{2N+1}jk},\, k=-N,\dots,N.$$

A simple calculation verifies that

$$g(k)=\frac{1}{\sqrt{2N+1}}\int \hat f(\xi)\hat\varphi(\xi)D_N\Big(\frac{T}{N}\xi-\frac{k}{2N+1}\Big);\qquad D_N(\xi)=\frac{\sin{(2N+1)\pi\xi}}{\sin\pi\xi}.$$

Thus the error between the samples of the Fourier transform f_φ and samples of the mollified Fourier transform $\hat f_\psi$ of f is quantified by

$$\widehat{f}_\psi\Big(\frac{\Omega k}{N}\Big)-g(k)=\int \hat f(\xi)\Big[\psi\Big(\frac{k\Omega}{N}-\xi\Big)-\frac{1}{\sqrt{2N+1}}\hat\varphi(\xi)D_N\Big(\frac{T}{N}\xi-\frac{k}{2N+1}\Big)\Big]\,d\xi.$$

Define the *error kernel*

$$\delta_{k,N,\Omega,T}(\xi)=\Big[\psi\Big(\frac{k\Omega}{N}-\xi\Big)-\frac{1}{\sqrt{2N+1}}\hat\varphi(\xi)D_N\Big(\frac{T}{N}\xi-\frac{k}{2N+1}\Big)\Big].$$

By Cauchy-Schwarz and Plancherel one has

$$\Big|\widehat{f}_\psi\Big(\frac{\Omega k}{N}\Big)-g(k)\Big|\le\|f\|\|\delta_{k,N,\Omega,T}\|$$

and the bound is sharp by taking f to be the error function $\delta_{k,N,\Omega,T}$ itself. Auslander and Grünbaum considered the case of Gaussian mollifiers

$$\varphi(t)=\frac{1}{\sqrt{2\pi\sigma_1}}\mathrm{e}^{-\frac{t^2}{2\sigma_1}};\qquad \psi(\xi)=\frac{1}{\sqrt{2\pi\sigma_2}}\mathrm{e}^{-\frac{\xi^2}{2\sigma_2}}$$

with the particular scalings

$$\sigma_1=c_1(T/N);\qquad \sigma_2=c_2(\Omega/N).$$

In this case it was shown that the *Wiener error* $\sum_{k=-N}^{N}\|\delta_{k,N,\Omega,T}\|$ is minimized at the *Nyquist rate* $T\Omega=(2N+1)/4$. This curious connection between Gaussian smoothing and uniform sampling merits further investigation.

The functions f_φ and $\widehat{f}_\psi$ are in the images of translation invariant systems. From the point of view of *joint* time-frequency localization translation it may not be natural to impose such invariance. One alternative is to replace f_φ by the projection of f onto the image of the eigenfunctions of $P_\Omega Q_T P_\Omega$ having large eigenvalues. That is, let

$$s_{\Omega,T,N}(t,u)=\sum_{k=1}^{N}\varphi_k(t)\varphi_k(u)$$

denote the kernel of the projection onto the first several eigenfunctions $\varphi_k=\varphi_k^{\Omega,T}$ of $P_\Omega Q_T P_\Omega$. Let $\psi_k=\psi_k^{T\Omega}$ be the corresponding eigenfunctions with the time and frequency roles reversed and denote by $\sigma_{T,\Omega,N}$ the corresponding kernel. In the special case $S=[-T/2,T/2]$ and $\Sigma=[-\Omega/2,\Omega/2]$ one can take advantage of the covariance of the eigenfunctions under the Fourier transform [18]. In particular, when normalized so that $\|\varphi_k\|=1$ and such that $T=\Omega=\sqrt{c}$ one has $|\widehat{\varphi_k}|=|\varphi_k|$ (see (5)) and then $\sigma_{T,\Omega,N}(\xi,\eta)=s_{\Omega,T,N}(\xi,\eta)$. Then if $f_N(t)=\int f(u)\,s_{\Omega,T,N}(t,u)\,du$ it follows that $\widehat{f_N}(\xi)=\sum\langle f,\varphi_k\rangle\widehat{\varphi_k}=(\widehat{f})_N(\xi)$. That is, the time-frequency projection commutes with the Fourier transform. Similar ideas were considered by Shen and Walter [41].

Problem 4 For f_N defined as above, quantify the dependence of the error between the DFT of the samples of f_N and the corresponding samples of $\widehat{f}_N$ on the sampling rate and the area $c = T\Omega$.

For general sets S and Σ there is no simple relationship between the eigenfunctions of $P_\Sigma Q_S P_\Sigma$ and their Fourier transforms, but one still has $\widehat{f_N}(\xi) = \sum_{k=1}^{N} \langle f, \varphi_k\rangle \widehat{\varphi_k}$. In any case it would be desirable to have a simple way to compute or estimate the coefficients $\langle f, \varphi_k\rangle$ in terms of the samples of f.

Problem 5 Express $\langle f, \varphi_k\rangle$, where φ_k are the eigenfunctions of $P_\Sigma Q_S P_\Sigma$, in terms of samples of f.

In the case $\Sigma = [-\Omega/2, \Omega/2]$ any f in the range of P_Σ can be expressed in terms of its samples $f(n/\Omega)$ and the coefficient $\langle f, \varphi_k\rangle$ can be written as $\frac{1}{\Omega}\sum_n f\big(\frac{n}{\Omega}\big)\varphi_k\big(\frac{n}{\Omega}\big)$, as will be discussed further in Section 6. A variant of Problem 5 will be stated later (as Problem 9) in the context of more general sampling schemes.

4.2 Discrete continuous functions

In addition to quantifying sampling errors, it is interesting to ask which families of functions in $L^2(\mathbb{R})$ can be characterized completely in terms of finite sets of samples, on one hand, and finitely many Fourier coefficients, on the other. Signals that are finite sums of shifted sinc functions are such examples, and their Fourier transforms can be expressed as trigonometric polynomials on their supports. It is of interest to consider broader classes of interpolating functions. Several such interpolation schemes are considered in Izu's dissertation [19], including extensions to distributions. Here we state one such interpolation result.

We will say that ψ is a T partition of unity if $\sum_{k=-\infty}^{\infty} \psi(t+kT) = 1$ for all $t \in \mathbb{R}$. We also say that a set S is a T-tiling set if the sets $\{kT + S\}_{k\in\mathbb{Z}}$ form a partition of $\mathbb{R}$.

Theorem 4 *Let $T\Omega \in \mathbb{N}$, let S be a T tiling set and let Σ be an Ω tiling set. Let $\psi \in \mathcal{S}(\mathbb{R})$ be a T partition of unity such that $\psi\big(\frac{n}{\Omega}\big) = 1$ if $n \in \Omega S$ and otherwise $\psi\big(\frac{n}{\Omega}\big) = 0$. Suppose that $\widehat{f}$ is a linear combination of shifts $\{\widehat{\psi}\big(\cdot - \frac{m}{T}\big) : m \in T\Sigma\}$. Then f satisfies*

$$\widehat{f}(\xi) = \frac{1}{T}\sum_{m\in T\Sigma} \widehat{f}\Big(\frac{m}{T}\Big)\widehat{\psi}\Big(\xi - \frac{m}{T}\Big) = \frac{1}{T\Omega}\sum_{n\in\Omega S} f\Big(\frac{n}{\Omega}\Big)\sum_{m\in T\Sigma} \mathrm{e}^{-2\pi i\frac{mn}{T\Omega}}\,\widehat{\psi}\Big(\xi - \frac{m}{T}\Big)$$

and

$$f(t) = \frac{1}{T}\sum_{m\in T\Sigma} \widehat{f}\Big(\frac{m}{T}\Big)\mathrm{e}^{2\pi i\frac{m}{T}t}\psi(t) = \frac{1}{T\Omega}\sum_{n\in\Omega S} f\Big(\frac{n}{\Omega}\Big)\sum_{m\in T\Sigma} \mathrm{e}^{-2\pi i\frac{m}{T}(\frac{n}{\Omega}-T)}\psi(t).$$

Example. Let $T = 36$ and $\Omega = 1/9$ so $T\Omega = 4$. Set $S = [0, 18) \cup [54, 72)$ and let $\Sigma = [0, 1/9)$. So $\{n \in \Omega S\} = \{0, 1, 6, 7\}$ while $\{m \in T\Sigma\} = \{0, 1, 2, 3\}$. Let $h(t) = \exp(-1/(1-|t|^2))$ for $|t| < 1$ and $h(t) = 0$ for $|t| > 1$. Let $H(t) = \sum_k h(t-k)$, let $g(t) = h(t)/H(t)$ and let

$$\psi(t) = \sum_{n\in\Omega S} g\big(\Omega t - n)\big) = g\Big(\frac{t}{9}\Big) + g\Big(\frac{t}{9} - 1\Big) + g\Big(\frac{t}{9} - 6\Big) + g\Big(\frac{t}{9} - 7\Big).$$

Then g is a 1-partition of unity. Also, ψ is a $T = 36$ partition of unity that vanishes at $9n$ if $n \notin \{0, 1, 6, 7\}$ since $h(k) = \delta_{0k}$. Finally, let

$$\widehat{f}(\omega) = \sum_{m \in T\Sigma} \alpha_m \widehat{\psi}\Big(\omega - \frac{m}{T}\Big) = \sum_{m=0}^{3} \alpha_m \widehat{\psi}\Big(\omega - \frac{m}{36}\Big).$$

Then f is as in the theorem so it can also be expressed in terms of its samples $\{f(0), f(9), f(54), f(63)\}$. One can then build functions f that can be recovered from samples along shifts by multiples of 72 of $\{0, 9, 54, 63\}$ by inverse Fourier transforming suitable sums of modulations of $\widehat{\psi}$.

4.3 Time frequency localization of multiband signals: some examples on $\mathbb{Z}_N$

One of the fundamental heuristics of time-frequency analysis is that one unit of area in the time-frequency plane corresponds to one degree of freedom. In the finite $N \times N$ time-frequency plane any *time-frequency tile of* N points corresponds roughly to one degree of freedom (e.g., [39]). We will define the *normalized area* of a product subset $S \times \Sigma \subset \mathbb{Z}_N \times \mathbb{Z}_N$ as $c = |S||\Sigma|/N$ where $|S|$ is the counting measure of S. By a $\mathbb{Z}_N$ interval we mean a nonempty subset $I \subset \mathbb{Z}_N$ that is maximal with respect to the property that if $k \in I$ then either $k+1 \in I$ or $k-1 \in I$ (with addition modulo N). Its endpoints are those $k \in I$ such that either $k-1 \notin I$ or $k+1 \notin I$. Intervals can be singletons. If $\Sigma \subset \mathbb{Z}_N$ we let N_Σ be the number of intervals in Σ.

Problem 6 Establish analogue(s) of the ΩT Theorem for $\mathbb{Z}_N$ (cf. [33]).

We have raised some questions about dependence of the spectrum of $P_\Sigma Q_S$ on the time and frequency supports S and Σ. The questions were posed for the continuous setting but numerical investigations, not to mention any digital implementations, have to be carried out in $\mathbb{Z}_N$. Here is a brief description of some numerical experiments concerning the dependence of the spectrum of $P_\Sigma Q_S P_\Sigma$ on the normalized time-frequency area c and on the distribution of the Fourier support Σ when S is a single $\mathbb{Z}_N$ interval. Our observations can be summarized briefly as follows. First, while $\lambda_{\lfloor c \rfloor}$ tends to be close to 1/2, there does not appear to be any rule more precise than the analogue of the Landau estimate ($\lambda_{\lfloor c \rfloor - 2N_\Sigma + 1} \geq 1/2 \geq \lambda_{\lceil c \rceil + 2N_\Sigma - 2}$). Secondly, the width of the *plunge region* over which the eigenvalues transition from being close to one to being close to zero appears to correlate with N_Σ.

We used centered DFTs of size $N = 2^K + 1$. Time intervals of length $T+1$ were also centered. Here is a description of the method used to produce frequency supports Σ. First we produced a random vector r of length $(N+1)/2 + L$ with values generated from a uniform distribution of values in $[0, 1]$. The parameter L serves to bias the intervals in Σ toward having length L. Specifically, we set $m(k) = \text{mean}[r(k), \ldots, r(k+L-1)]$, $k = 1, \ldots, (N+1)/2$, and took $k \in \Sigma$ if $m(k)$ was among the largest $|\Sigma|/2$ values of m. In this way the normalized area could be fixed. Finally we symmetrized: for $k = (N+3)/2, \ldots, N$ we let $k \in \Sigma$ if $N - k \in \Sigma$. We did not keep track of the lengths of the individual intervals although we did keep track of the number N_Σ of symmetric intervals. Numerical results are given in Tables 1 – 3 and Figs. 1 and 3.

Some additional comments can be made, keeping in mind that they are based on a small sample and also on a small N. First, there is significant variation in the eigenvalues for fixed c and L, indicating a significant dependence on the distribution

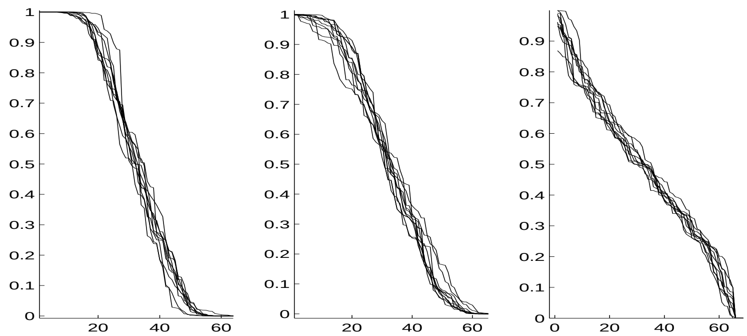

Fig. 1 *Eigenvalue distributions* Eigenvalue distributions for randomly generated frequency supports, $N = 257$, $T = 64$, $N_\Sigma = 128$. The figure on the left shows the eigenvalue distribution for ten randomly generated Σ's with $L = 10$, biased toward longer intervals. The middle plot shows similar distribution with $L = 5$ and the plot on the right is similar with $L = 1$, biased toward shorter intervals. The normalized area is $c = 32$ in all cases. The eigenvalue curves change from being convex for large L to being concave for small L.

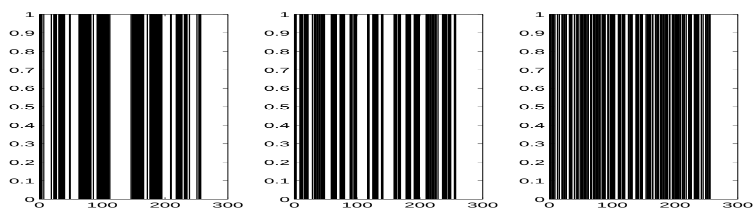

Fig. 2 *Frequency supports* Typical frequency supports Σ, $|\Sigma| = 128$ for $L = 10$ (left), $L = 5$ (middle) and $L = 1$ (right).

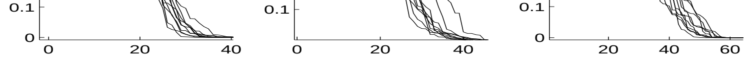

Fig. 3 *Eigenvalue distributions* Eigenvalue distributions for randomly generated frequency supports, $N = 257$, $T = 64$, $N_\Sigma = 64$. The frequency sets are half the size of the ones in Fig. 1. As in that figure, $L = 10$ (left), $L = 5$ (middle) and $L = 1$ (right). The normalized area is $c = 16$ in all cases. When $L = 1$ the largest eigenvalues are no longer close to one while the small but nontrivial eigenvalues extend well beyond $2c$ in that case.

Table 1 Eigenvalues of $P_\Sigma Q_T$ with random frequency supports, $N = 257$, $T = 64$, $|\Sigma| = 128$, $c = 32$, $L = 10$. Average value of $\lambda_{\lfloor c \rfloor}$ is 0.52.

trial #	1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
N_Σ	12	14	10	9	13	13	7	8	13	8
$\lambda_{\lfloor c \rfloor -3}$	.57	.57	.61	.63	.59	.59	.66	.69	.57	.65
$\lambda_{\lfloor c \rfloor -2}$	.57	.57	.58	.62	.59	.56	.66	.64	.56	.65
$\lambda_{\lfloor c \rfloor -1}$	.53	.57	.56	.60	.58	.56	.62	.59	.55	.60
$\lambda_{\lfloor c \rfloor}$	.49	.54	.52	.48	.50	.55	.61	.50	.52	.53
$\lambda_{\lfloor c \rfloor +1}$	.46	.52	.43	.48	.50	.55	.57	.43	.49	.52
$\lambda_{\lfloor c \rfloor +2}$	.43	.50	.42	.43	.50	.52	.51	.42	.49	.51
$\lambda_{\lfloor c \rfloor +3}$	.39	.42	.35	.41	.48	.49	.44	.33	.46	.47

Table 2 Eigenvalues of $P_\Sigma Q_T$ with random frequency supports, , $N = 257$, $T = 64$, $|\Sigma| = 128$, $c = 32$, $L = 5$. Average value of $\lambda_{\lfloor c \rfloor}$ is 0.55

trial #	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
N_Σ	15	14	13	16	14	15	10	15	9	10
$\lambda_{\lfloor c \rfloor -3}$	.62	.59	.66	.61	.65	.58	.65	.66	.68	.71
$\lambda_{\lfloor c \rfloor -2}$	.61	.55	.63	.57	.61	.58	.65	.60	.59	.67
$\lambda_{\lfloor c \rfloor -1}$	.60	.54	.60	.55	.59	.51	.61	.59	.58	.67
$\lambda_{\lfloor c \rfloor}$	.52	.53	.59	.53	.55	.49	.60	.53	.53	.58
$\lambda_{\lfloor c \rfloor +1}$	.51	.50	.59	.51	.48	.49	.57	.51	.52	.53
$\lambda_{\lfloor c \rfloor +2}$	.45	.48	.50	.49	.45	.44	.56	.48	.43	.46
$\lambda_{\lfloor c \rfloor +3}$	.43	.47	.50	.45	.44	.39	.50	.44	.39	.39

Table 3 Eigenvalues of $P_\Sigma Q_T$ with random frequency supports, , $N = 257$, $T = 64$, $|\Sigma| = 128$, $c = 32$, $L = 1$. Average value of $\lambda_{\lfloor c \rfloor}$ is 0.51

trial #	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
N_Σ	27	35	34	30	30	33	30	32	30	32
$\lambda_{\lfloor c \rfloor -3}$	.58	.50	.56	.53	.54	.56	.53	.54	.57	.56
$\lambda_{\lfloor c \rfloor -2}$	.58	.50	.55	.50	.52	.56	.52	.52	.56	.55
$\lambda_{\lfloor c \rfloor -1}$	.57	.50	.53	.50	.50	.54	.49	.51	.55	.50
$\lambda_{\lfloor c \rfloor}$	.56	.49	.53	.49	.50	.51	.49	.51	.53	.50
$\lambda_{\lfloor c \rfloor +1}$	.55	.46	.52	.49	.46	.51	.49	.50	.51	.50
$\lambda_{\lfloor c \rfloor +2}$	.54	.46	.51	.48	.46	.51	.48	.50	.50	.50
$\lambda_{\lfloor c \rfloor +3}$	.53	.46	.49	.45	.46	.50	.48	.49	.49	.48

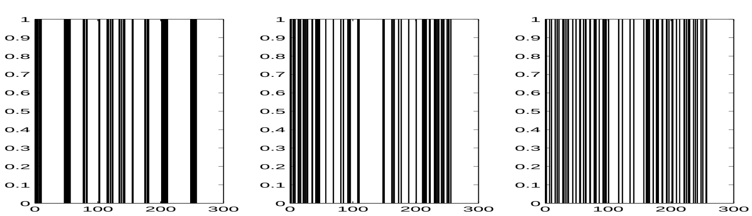

Fig. 4 *Frequency supports* Typical frequency supports Σ, $|\Sigma| = 64$, for $L = 10$ (left), $L = 5$ (middle) and $L = 1$ (right).

of Σ. In addition, there appears to be a general trend toward more eigenvalues less than $1/2$ as Σ becomes more disconnected. As c becomes smaller a larger proportion of trace norm $T|\Sigma|$ appears to be accounted for by the eigenvalues less than $1/2$.

5 Finite case: sparsity and connection with compressive sampling

A fundamental difference between *finite* and *infinite* settings is manifested in support properties of Fourier transforms. No nonzero $f \in L^2(\mathbb{R})$ that vanishes outside a set of finite measure has a Fourier transform $\hat{f}$ also vanishing outside a set of finite measure. The analogous statement for the DFT is a lower bound on the sizes of the supports of $\mathbf{x}$ and $\widehat{\mathbf{x}}$ due to Donoho and Stark [10].

The Donoho-Stark inequality. As before, $|F|$ denotes the counting measure of a finite set F. The Donoho-Stark uncertainty inequality can be stated as follows.

Theorem 5 *Any finite signal $\mathbf{x} : \mathbb{Z}_N \to \mathbb{C}$ with N-point discrete Fourier transform $\widehat{\mathbf{x}}$ satisfies*

$$|\{n : x_n \neq 0\}|\,|\{n : \widehat{x}_n \neq 0\}| \geq N, \tag{10}$$

$$|\{n : x_n \neq 0\}| + |\{n : \widehat{x}_n \neq 0\}| \geq 2\sqrt{N}. \tag{11}$$

In addition, $|\{n : x_n \neq 0\}|\,|\{n : \widehat{x}_n \neq 0\}| = N$ only when, up to appropriate time-frequency shifts, $\mathbf{x}$ is the indicator of a subgroup where $\mathbb{Z}_N$ is regarded as a cyclic group of order N.

The characterization of extremal functions for the Donoho-Stark uncertainty emphasizes an important (group theoretic) distinction between the DFT and the Fourier transform on $\mathbb{R}$. The role of group theory is emphasized further in Tao's [38] sharper inequality for $N = P$ prime, namely

$$|\{n : x_n \neq 0\}| + |\{n : \widehat{x}_n \neq 0\}| \geq P + 1.$$

Recall that $\mathbb{Z}_P$ has no nontrivial proper subgroups. More information regarding the Donoho-Stark inequality can be found in [18].

Quantitative robust uncertainty principles. The finite setting can indicate limitations of the multiband cases of the Bell Labs time- and bandlimiting theory on $\mathbb{R}$ – not to mention precise limitations of the analogous finite theory – when the time or frequency supports become highly disconnected. Candès, Romberg and Tao [9, 7] found that norm estimates in such disconnected cases could be useful in signal recovery problems. Specifically, they considered the problem of finding a bound on the norm – and hence on the largest eigenvalue – of the discrete version of the operator $P_\Sigma Q_S P_\Sigma$ when the time-frequency area is small. Recall that in the finite case the normalized area is $|S||\Sigma|/N$ where $|S|$ is the counting measure of S. Denote by $A_{S\Sigma}$ the operator with standard basis matrix $D_S \mathcal{F}_N^{-1} D_\Sigma \mathcal{F}_N$ where $\mathcal{F}_N$ is the matrix of the N-point DFT and $D_S = \operatorname{diag} S$ is the diagonal matrix with $D_S(j, j) = 1$ if $j \in S$ and $D_S(j, k) = 0$

otherwise. Thus D_S is the matrix of multiplication by the discrete indicator function $\mathbb{1}_S$. Then

$$\begin{aligned} A_{S\Sigma}A_{S\Sigma}^* &= D_S\mathcal{F}_N^{-1}D_\Sigma\mathcal{F}_N(D_S\mathcal{F}_N^{-1}D_\Sigma\mathcal{F}_N)^* \\ &= D_S\mathcal{F}_N^{-1}D_\Sigma\mathcal{F}_N\mathcal{F}_N^*D_\Sigma\mathcal{F}_N D_S = D_S\mathcal{F}_N^{-1}D_\Sigma\mathcal{F}_N D_S \end{aligned} \tag{12}$$

since $\mathcal{F}_N^* = \mathcal{F}_N^{-1}$ and D_Σ is idempotent.

Candès, Romberg and Tao (CRT) were motivated by applications to compressed sensing in which a signal $\mathbf{x}$ could be recovered (using nonlinear optimization techniques) from its values on S provided that its Fourier transform $\widehat{\mathbf{x}}$ vanishes outside Σ. This recovery is contingent upon invertibility of $I - A_{S\Sigma}$, which holds if $\|A_{S\Sigma}\| \ll 1$. Candès, Romberg and Tao were able to obtain such bounds in a probabilistic sense. We will give a brief outline of their approach here, and a more detailed outline of their technical estimates in Appendix A. For technical reasons, CRT assume $N \geq 512$. They fix a parameter $1 \leq \beta \leq (3/8)\log N$. Set

$$M(N,\beta) = \frac{N}{\sqrt{(\beta+1)\log N}}\Big(\frac{1}{\sqrt{6}} + o(1)\Big).$$

The "$o(1)$" term arises through Proposition 4 and Lemma 2 below.

Theorem 6 *Suppose that S and Σ are subsets of $\mathbb{Z}_N$ whose sizes satisfy $|S| + |\Sigma| \leq M(N,\beta)$. Then with probability at least $1 - O((\log N)^{1/2}/N^\beta)$, every signal $\mathbf{x}$ supported in S satisfies*

$$\|\widehat{\mathbf{x}}\mathbb{1}_\Sigma\|^2 \leq \frac{1}{2}\|\mathbf{x}\|^2$$

while every signal $\mathbf{x}$ frequency supported in Σ satisfies

$$\|\mathbf{x}\mathbb{1}_S\|^2 \leq \frac{1}{2}\|\mathbf{x}\|^2.$$

The second inequality says that $\|A_{S\Sigma}\|^2 \leq 1/2$. By the arithmetic-geometric inequality,

$$\frac{|S||\Sigma|}{N} \leq \frac{1}{4N}M(N,\beta)^2 \leq \frac{N}{24(\beta+1)\log N}(1+o(1)).$$

This suggests that, even for *small* N, there can exist time-frequency set pairs of normalized area substantially larger than one on which no signal is mostly localized.

The probability is defined with respect to the uniform distribution among all sets S and Σ of fixed sizes $|S|$ and $|\Sigma|$. There are two main ingredients to the proof. The CRT methods require ability to treat the frequency coordinates independently. For this reason they considered Bernoulli generated frequency sets of random size rather than sets of fixed size. Thus the first ingredient of the proof involves the random choice of frequency support. In contrast to our earlier usage of the symbol Ω as an interval length, we will use the symbol Ω here to denote a *Bernoulli* generated frequency support set of random size $|\Omega|$, while Σ denotes a *uniformly* generated support set of fixed size $|\Sigma|$. To describe Ω consider, for each $\omega \in \mathbb{Z}_N$, a Bernoulli random variable $I(\omega)$ (that is, a random variable with values in $\{0,1\}$) with $\mathrm{prob}\,(I(\omega)=1) = \tau$ for each ω. Set $\Omega = \{\omega \in \mathbb{Z}_N : I(\omega) = 1\}$. Then $|\Omega| = \sum_{\omega=0}^{N-1} I(\omega)$ has expected value $N\tau$. Take $\tau = |\Sigma|/N$ where, as before, $|\Sigma|$ is fixed. One has to show that $|\Omega|$ does not vary too much from its expected value of $|\Sigma|$. The second ingredient is a probabilistic bound on the norm of $A_{S\Omega}$ with S fixed.

The main technical estimate for Theorem 6. Consider now the auxiliary matrix

$$H_{S\Omega}(s,t) = \begin{cases} \text{diag} \sum_{\omega\in\Omega} \mathrm{e}^{2\pi i\omega(s-t)/N}, & s,t\in S, s\neq t \\ 0, & \text{else} \end{cases}. \tag{13}$$

Then $A_{S\Omega}A_{S\Omega}^* = \frac{|\Omega|}{N}I + \frac{1}{N}H_{S\Omega}$ and

$$\|A_{S\Omega}\|^2 \leq \frac{|\Omega|}{N} + \frac{\|H_{S\Omega}\|}{N}. \tag{14}$$

The CRT technique for estimating $\|H_{S\Omega}\|$ is to estimate traces of powers of $H_{S\Omega}^2$, similarly to Landau and Widom's approach to proving Theorem 3. We include a detailed outline of the proof of the following in Appendix A since the sophisticated techniques should be useful in addressing some of the open problems mentioned below. The main estimate is the following.

Proposition 3 *If $\tau \leq \mathrm{e}^{-2}$ then the powers of $H_{S\Omega}$ satisfy the expectation inequality*

$$\mathbf{E}(\mathrm{tr}(H_{S\Omega}^{2n})) \leq 2\,n^{n+1}\left(\frac{4}{(1-\tau)\mathrm{e}}\right)^n |S|^{n+1}(\tau N)^n.$$

Bernoulli versus uniform models. Theorem 6 states that for $|S|+|\Sigma|$ *small*, $\|A_{S\Sigma}\| \leq 1/2$ with overwhelming probability, taken with respect to the uniform distribution of sets of size $|\Sigma|$. But Proposition 3 applies to Bernoulli generated sets Ω whose sizes are also random. To pass from the proposition to the theorem one needs probabilistic bounds relating the probabilities of the operators $A_{S\Sigma}$ and $A_{S\Omega}$ having large norms. Such a bound was obtained by Càndes and Romberg [7] as follows.

Lemma 1 *Let Σ be drawn uniformly at random from the subsets of $\mathbb{Z}_N$ of fixed size $|\Sigma|$ with $(\log N)^2 \leq |\Sigma| \leq N/\sqrt{6\log N}$, and let $\Omega = \{\omega\in\mathbb{Z}_N : I(\omega)=1\}$ where, for each ω,* $\mathrm{prob}\,(I(\omega)=1) = |\Sigma|/N$*. Let $A_\Omega = A_{S\Omega}$ and $A_\Sigma = A_{S\Sigma}$ with the same fixed time support S. Then*

$$\mathrm{prob}\,(\|A_\Omega\|^2 > 1/2) \geq \frac{1}{2}\mathrm{prob}\,(\|A_\Sigma\|^2 > 1/2).$$

This lemma follows essentially from the monotonicity of $\|A_\Omega\|$, that is, $\|A_{\Omega_1}\| \leq \|A_{\Omega_2}\|$ if $\Omega_1 \subset \Omega_2$, the fact that for Bernoulli generated Ω, if $\mathbf{E}(|\Omega|)\in\mathbb{N}$ then the median and expected value of $|\Omega|$ are equal, and the fact that for the given bounds on $|\Sigma|$, $|\Omega|$ is not expected to deviate from its mean too much.

Probabilistic norm estimates for the Bernoulli model. Armed with Lemma 1, Theorem 6 reduces to a corresponding estimate for the Bernoulli case. In other words, with $S\subset\mathbb{Z}_N$ a fixed time support with $|S|$ and ρ, q and β as before one has the following.

Proposition 4 *Let $\rho = \sqrt{2/3}$ and fix $1 \leq \beta \leq (3/8)\log N$ and $q\in(0,1/2]$. Let $\Omega = \{\omega\in\mathbb{Z}_N : I(\omega)=1\}$ where* $\mathrm{prob}\,(I(\omega)=1) = |\Sigma|/N$ *and*

$$|S|+|\Sigma| \leq \frac{\rho q N}{\sqrt{(\beta+1)\log N}}.$$

Let $H_{S\Omega}$ be defined as in (13). Then

$$\mathrm{prob}\,(\|H_{S\Omega}\| > qN) \leq 5N^{-\beta}\sqrt{(\beta+1)\log N}.$$

The moment bound of Proposition 3 is utilized as follows. Since $H = H_{S\Omega}$ is self-adjoint the Markov inequality and $\|H^n\| \leq \mathrm{tr}(H^n)$ implies that for $n = 1, 2, \ldots,$

$$\mathrm{prob}\,(\|H\| > qN) = \mathrm{prob}\,(\|H^n\|^2 > (qN)^{2n}) \leq \frac{\mathbf{E}\|H^n\|^2}{(qN)^{2n}} \leq \frac{\mathbf{E}(\mathrm{tr}(H^{2n}))}{(qN)^{2n}}. \tag{15}$$

Taking $\tau < 1/3$ if necessary (so that $4/(1-\tau) \leq 6$) one obtains from Proposition 3 and the inequality on arithmetic-geometric means that

$$\mathbf{E}(\mathrm{tr}(H^{2n})) \leq 2n(6/\mathrm{e})^n n^n |S|^{n+1} (\tau N)^n \leq 2\frac{n^{n+1}}{\mathrm{e}^n}|S|\left(\frac{|S| + \tau N}{\rho}\right)^{2n}.$$

Specializing to $n = \lfloor(\beta+1)\log N\rfloor$, the hypothesis $|S| + \tau N \leq \frac{\rho q N}{\sqrt{(\beta+1)\log N}}$ and (15) finally imply that

$$\mathrm{prob}\,(\|H\| > qN) \leq 2\mathrm{e}\rho q N^{-\beta}\sqrt{(\beta+1)\log N} \leq 5N^{-\beta}\sqrt{(\beta+1)\log N}.$$

This establishes Proposition 4.

To pass from a probabilistic bound on $H_{S\Omega}$ to one on $A_{S\Omega}$, see (14), one requires that $|\Omega|$ does not deviate too much. When $\tau \geq (\log N)^2/N$, Ω is likely not too large as the following lemma shows.

Lemma 2 *With ρ, β, q as in Proposition 4, let $\Omega = \{\omega \in \mathbb{Z}_N : I(\omega) = 1\}$ where $\mathbf{E}(|\Omega|) = \tau N$ with*

$$\frac{(\log N)^2}{N} \leq \tau \leq \frac{\rho q}{\sqrt{(1+\beta)\log N}}.$$

Then, with probability at least $1 - N^{-\beta}$,

$$\frac{|\Omega|}{N} \leq \frac{2\rho q}{\sqrt{(1+\beta)\log N}}.$$

Thus, with probability at least $1 - 0(N^{-\beta}\sqrt{\log N})$,

$$\|A_{S\Omega}\|^2 \leq q\left(1 + \frac{2\rho}{\sqrt{(\beta+1)\log N}}\right).$$

Choosing $q = \frac{1}{2}(1 - (2\rho/\sqrt{(\beta+1)\log N})) = 1/2 + o(N)$ proves that $\|A_{S\Omega}\|^2 < 1/2$ with the desired probability. Theorem 6 then follows from Lemma 1.

CRT in action. As an example, Candès, Romberg and Tao consider $|S| + |\Sigma| \leq \frac{0.2971N}{\sqrt{(\beta+1)\log N}}$. For simplicity we then take $\beta = (13/36)\log N < (3/8)\log N$ which places the bound $|S| + |\Sigma| \leq \frac{0.2971N}{(7/6)\log N} = 0.2547N/\log N$. When $N = 512$ this gives $|S| + |\Sigma| \leq 20.88$. Theorem 6 says that under this constraint the probability that there is an $\mathbf{x}$ frequency supported on Σ at least half of whose energy lies in S is $O(\log(N)^{1/2}/N^{\beta})$ with an undetermined constant C. In our case, the probability is at most $C\sqrt{6.24}/(512)^{13/36} \leq C(2.5)/(2^{13/4}) \approx (1/4)$.

As an experiment we generated 100 time and frequency support sets of the same size ($|S| = |\Sigma|$). The sets were generated by taking the indices of the largest $|S|$ values of a vector of dimension N with coordinates generated by a uniform random

Table 4 Norm of $A = \mathcal{F}^{-1}\mathbb{1}_\Sigma \mathcal{F}\mathbb{1}_S$ for randomly generated supports, $N = 512$

$\|S\| = \|\Sigma\|$	c	max $\|A\|$	min $\|A\|$	$\%\|A\|^2 > 1/2$
10	0.20	.27	.22	0
20	0.78	0.39	0.33	0
40	3.12	0.54	0.49	0
80	12.5	0.73	0.69	59
100	19.53	0.80	0.76	100

Table 5 Norm of $A = \mathcal{F}^{-1}\mathbb{1}_\Sigma \mathcal{F}\mathbb{1}_S$ for centered time support and randomly generated frequency support, $N = 512$

$\|S\| = \|\Sigma\|$	c	max $\|A\|$	min $\|A\|$	$\%\|A\|^2 > 1/2$
10	0.20	0.29	0.20	0
20	0.78	0.44	0.30	0
40	3.13	0.61	0.44	0
80	12.5	0.88	0.66	70
100	19.53	0.93	0.73	100

Table 6 Norm of $A = \mathcal{F}^{-1}\mathbb{1}_\Sigma \mathcal{F}\mathbb{1}_S$ for centered time support and frequency supports $N = 512$

$\|S\| = \|\Sigma\|$	c	$\|A\|$
10	0.20	0.44
20	0.78	0.82
40	3.13	0.88
80	12.5	0.9996
100	19.53	1.0000

distribution. With $N = 512$ we report the maximum and minimum of the norms of $A_{S\Sigma} = \mathcal{F}^{-1}D_\Sigma \mathcal{F} D_S$.

Table 4 lists norms of the matrix $A = \mathcal{F}^{-1}D_\Sigma \mathcal{F} D_S$ when S and Σ are both uniformly randomly generated. The supports are obtained by choosing the largest $M = |S| = |\Sigma|$ elements from N elements randomly generated from a uniform $[0, 1]$ distribution. For $N = 512$ the normalized time frequency area c needs to be approximately 10 before one starts finding $\mathbf{x}$ frequency supported in Σ and having at least half its energy in S.

Table 5 shows the eigenvalues in the case in which the time interval is centered at the origin but the frequency support is randomly generated. For comparison, Table 6 illustrates the growth in the norm of A when the time and frequency supports are both "intervals" centered around $\lfloor N + 1\rfloor/2$. In this case one gets a norm bigger than $1/2$ even when the time-frequency area is strictly less than unity.

We also considered the question of how the norm of A scales with N. Here are some brief observations. When $N = 1024$, and $|S| = |\Sigma| = 112$, corresponding to $c \approx 12.5$, the norm of A ranges from 0.5964 to 0.6209. For $N = 2048$ and $|S| = |\Sigma| = 160$, again corresponding to $c \approx 12.5$, the norm of A ranges from 0.5179 to 0.5308. For $N = 1024$ and $|S| = |\Sigma| = 14$ corresponding to $c = 0.195$ the norm of A ranges from 0.1933 to 0.2119, taking 10 samples in each case. In particular, for a given normalized time-frequency area, the expected norm of A appears to decrease with N. This decay is consistent with the appearance of the factor $1/N^\beta$ in Theorem 6. Its combinatorial basis, in a nutshell, is that as N grows there is a larger proportion of sets Σ all of whose frequencies necessarily interfere destructively substantially over S.

When are the CRT estimates useful? The quantitative, probabilistic uncertainty principle estimates of Candès, Romberg and Tao were motivated by *compressive sampling* – the possibility of recovering a signal having a sparse but otherwise unknown Fourier spectrum from a small number of measurements, with a high probability of success. The methods are extremely important in such situations. In many radio frequency (RF) applications, on the other hand, the full RF spectrum is not sparse and is even dense in some regions, containing several frequency bands in which energy is localized. Frequency allocation charts for the United States can be found on the National Telecommunications and Information Administration web pages or the Federal Communications Commission web pages. It makes sense to ask whether the reasoning behind the CRT estimates applies to such a setting or to a heavily filtered version thereof. For a given normalized time-frequency area, the vast majority of cases giving small norms for $A_{S\Sigma}$ in the CRT estimates involve highly disconnected Fourier supports. This needs to be quantified more precisely.

Spectrum entropy. If Σ is a finite union of intervals, it induces a partition of its convex hull $[a, b]$ into intervals. One can then define the entropy of Σ as the entropy of the associated partition. Recall that the entropy of a partition $\mathcal{P}$ of a probability space $(X, \mathcal{B}, \mu)$ is $E(P) = -\sum_{P\in\mathcal{P}} \mu(P) \log \mu(P)$. For each interval P we can set $\mu(P) = |P|/(b-a)$.

Problem 7 Let $S = [-T/2, T/2]$. For fixed $|\Sigma|$ and hence fixed time-frequency area, establish a quantitive, probabilistic relationship between the entropy of Σ and the norm of $P_\Sigma Q_T$ in the continuous and or finite settings.

The partition entropy is not a perfect measure of disorder in all cases. For example, entropy can be large if Σ consists of a large number of evenly spaced intervals even though, in light of the Donoho-Stark uncertainty principle, this would be a case that allows $P_\Sigma Q_T$ or at least its finite analogue to have large norm. This is one reason for suggesting a probabilistic approach. It is of particular interest whether the probability that $P_\Sigma Q_T$ has one or several large eigenvalues decreases gradually as Σ becomes more disconnected or if this probability exhibits a sort of phase transition – a sharp decrease as entropy or some other measure of disorder increases.

6 Sampling, interpolation, and eigenfunctions

As previously, suppose that $S \subset \mathbb{R}$ and $\Sigma \subset \mathbb{R}$ are compact. Suppose that $\{x_n\} \subset \mathbb{R}$ and $\{\psi_n\} \subset \mathrm{PW}_\Sigma$ have the *sampling/interpolation* property

$$f(t) = \sum_n f(x_n)\psi_n(t), \qquad (f \in \mathrm{PW}_\Sigma). \tag{16}$$

To each sample point x_n associate $\rho_n(t) \equiv \big(\mathbb{1}_\Sigma\big)^\vee (x_n - t)$ and define a matrix

$$A : A_{nm} = \int_S \rho_n(t)\psi_m(t)\, dt. \tag{17}$$

Since $P_\Sigma g = g * (\mathbb{1}_\Sigma)^\vee$, on PW_Σ one has the *sampling formula*

$$P_\Sigma Q_S f(x_n) = \int_S \Big(\sum_m f(x_m)\psi_m(t)\Big)\rho_n(t)\,dt = \sum_m A_{nm} f(x_m) \tag{18}$$

and the *interpolation formula*

$$P_\Sigma Q_S f(t) = \sum_n (P_\Sigma Q_S f)(x_n)\,\psi_n(t) = \sum_n \Big(\sum_m A_{nm} f(x_m)\Big)\psi_n(t) \tag{19}$$

As in [19] one then has the following.

Theorem 7 *If φ is a λ-eigenfunction of $P_\Sigma Q_S$ then $\{\varphi(x_n)\}$ is a λ-eigenvector of A. Conversely, if $\mathbf{v}$ is a λ-eigenvector of A and if $\varphi(t) = \sum_m v_m \psi_m(t)$ converges then φ is a λ-eigenfunction of $P_\Sigma Q_S$.*

Proof Suppose that φ is a λ-eigenfunction. Then (18) implies

$$\lambda\varphi(x_n) = (P_\Sigma Q_S \varphi)(x_n) = \sum_m A_{nm}\varphi(x_m),$$

that is, $\{\varphi(x_n)\}$ is a λ-eigenvector of A.

To prove the converse we observe first that (16) implies $\psi_n(x_m) = \delta_{nm}$ so if $\varphi(t) = \sum_m v_m\psi_m(t)$ converges then $\varphi(x_m) = v_m$. If $\mathbf{v}$ is an eigenvector of A then (19) gives

$$\begin{aligned} P_\Sigma Q_S\varphi(t) &= \sum_n \Big(\sum_m A_{nm}\varphi(x_m)\Big)\psi_n(t) \\ &= \sum_n \Big(\sum_m A_{nm} v_m\Big)\psi_n(t) = \lambda \sum_n v_n \psi_n(t) = \lambda\varphi(t). \end{aligned} \tag{20}$$

That is, $\varphi(t) = \sum_n v_n\psi_n(t)$ is a λ-eigenfunction.

The case of PSWFs. The sampling theorem tells us that any $f \in \mathrm{PW}_\Omega$ satisfies

$$f(x) = \frac{1}{\Omega}\sum_{k\in\mathbb{Z}} f\Big(\frac{k}{\Omega}\Big)\mathrm{sinc}_\Omega\Big(x - \frac{k}{\Omega}\Big).$$

In the terms above one has $x_k = k/\Omega$ and $\rho_k(x) = \frac{1}{\Omega}\mathrm{sinc}_\Omega\big(x - \frac{k}{\Omega}\big) = \psi_k(x)$. In this case the sampling formula (18) becomes

$$\lambda_n\varphi_n\Big(\frac{m}{\Omega}\Big) = \sum_{k\in\mathbb{Z}} A_{mk}\varphi_n\Big(\frac{k}{\Omega}\Big)$$

where

$$A_{mk} = \frac{1}{\Omega^2}\int_{-T/2}^{T/2} \mathrm{sinc}_\Omega\Big(x - \frac{m}{\Omega}\Big)\mathrm{sinc}_\Omega\Big(x - \frac{k}{\Omega}\Big)\,dx. \tag{21}$$

Here the matrix A is real and symmetric, hence has a decomposition $A = UDU^T$ in which D contains is diagonal and the columns of U are the sample vectors $\varphi_n(k/\Omega)$ and one has the orthogonality

$$\frac{1}{\Omega}\sum_{k\in\mathbb{Z}} \varphi_n\Big(\frac{k}{\Omega}\Big)\varphi_\ell\Big(\frac{k}{\Omega}\Big) = \delta_{n,\ell}, \tag{22}$$

stating that the sequences of samples of the $\{\varphi_n\}$ form an orthonormal basis for the space $\mathrm{SPW}_\Omega \cong \ell^2(\mathbb{Z})$ of samples of PW_Ω. Khare and George [21] and Shen and Walter [41] showed this independently and they also showed that

$$\sum_{n=0}^{\infty} \varphi_n\Big(\frac{m}{\Omega}\Big)\varphi_n\Big(\frac{\ell}{\Omega}\Big) = \Omega\delta_{m,\ell}, \tag{23}$$

which says that the sum over n of the PSWF samples forms the reproducing kernel (the Dirac delta) for the sample space. An interesting problem, which we state imprecisely here, is to quantify the extent to which these identities approximately hold for the space of essentially time- and bandlimited signals.

Problem 8 Establish asymptotic error bounds for the quantities

$$E(n,\ell,A_1) = \sum_{m=-A_1}^{A_1} \varphi_n\Big(\frac{m}{\Omega}\Big)\varphi_\ell\Big(\frac{m}{\Omega}\Big) \quad \text{and} \quad F(m,\ell,A_2) = \sum_{n=0}^{A_2} \varphi_n\Big(\frac{m}{\Omega}\Big)\varphi_n\Big(\frac{\ell}{\Omega}\Big)$$

with A_1 and A_2 regarded in terms of the area ΩT.

Double orthogonality revisited. In Section 2 we saw that the eigenfunctions of $P_\Sigma Q_S$ are orthogonal over S – as well as over $\mathbb{R}$ – provided the kernel of $P_\Sigma Q_S$ is real-valued and even. In the case of the PSWFs this double orthogonality also follows from (22), since

$$\begin{aligned}
\int_{-T/2}^{T/2} \varphi_k(x)\varphi_\ell(x)\,dx &= \sum_{m\in\mathbb{Z}} \varphi_k\Big(\frac{m}{\Omega}\Big) \int_{-T/2}^{T/2} \operatorname{sinc}_\Omega\Big(x - \frac{m}{\Omega}\Big)\,\varphi_\ell(x)\,dx \\
&= \lambda_\ell \sum_{m\in\mathbb{Z}} \varphi_k\Big(\frac{m}{\Omega}\Big)\varphi_\ell\Big(\frac{m}{\Omega}\Big) \\
&= \Omega\lambda_\ell\delta_{k,\ell}.
\end{aligned}$$

Projections and sampling. The Shannon sampling theorem says that any $f \in \mathrm{PW}_\Omega$ can be recovered from its samples along $\mathbb{Z}/\Omega$. The orthogonality of the eigenvectors $\{\varphi_n(m/\Omega)\}$ of A in (21) allows one to compute the projection of $f \in \mathrm{PW}_\Omega$ onto the span of the first several eigenfunctions of $P_\Omega Q_T$ in a particularly convenient way because one only needs to compute the sample inner products. On the other hand, pointwise decay of the eigenfunctions suggests that the projection can be computed solely from those samples near $[-T/2, T/2]$ to within a given error with estimates parallel to those sought in Problem 8.

This also raises the question of what is required in order to extend this *samples to projections* method to the case of $P_\Sigma Q_S$ operators. Obviously one wants first that the conditions of Theorem 7 hold.

Problem 9 Suppose that $\{x_n\} \subset \mathbb{R}$ and $\{\psi_n\} \subset \mathrm{PW}_\Sigma$ have the sampling and interpolation property (16) and let the matrix A be as in (17) so that the eigenvectors of A correspond to the eigenfunctions of $P_\Sigma Q_S$ as in Theorem 7. Formulate the orthogonal projection onto the direct sum of the eigenspaces of the N largest eigenvalues as a mapping from the samples $\{f(x_n)\}$ to the coefficients $\langle f, \varphi_k\rangle$ of f with the corresponding eigenvectors of $P_\Sigma Q_S$. In addition, provide quantitative estimates for the errors that arise when only samples *near* S are used.

Of course this problem cannot be addressed without understanding first the nature of the interpolating functions ψ_n in (16). In the next section we will review some methods for their construction in the case of multiband signals.

7 Sampling of multiband signals

Theorem 7 presumes the existence of interpolating functions ψ_n such that any $f \in \mathrm{PW}_\Sigma$ can be written $f(t) = \sum_n f(x_n)\psi_n(t)$. Up to this point little has been said about how one might construct such interpolating functions. When Σ is a finite, pairwise disjoint union of intervals a number of sampling methods have been developed. We will concentrate on two such methods, those due to Venkataramani and Bresler [40] and to Behmard and Faridani [3] (cf. also [4] for more recent work). Two notable related approaches that rely on iterated filterbanks are due to Herley and Wong [13] and to Eldar and Oppenheim [12]. These approaches and others conceptually lie somewhere between those of Venkataramani and Bresler and of Behmard and Faridani.

7.1 Multicoset approach (Venkataramani and Bresler)

The work of Venkataramani and Bresler identifies interpolating functions for *periodic nonuniform sampling* or *multicoset sampling* of multiband signals. Let $\Sigma = \cup_{k=1}^N [a_k, b_k]$ where $b_k < a_{k+1}$ and, for convenience, let $a_1 = 0$ and $b_N = 1/T$. Then $\mathrm{PW}_\Sigma \subset \mathrm{PW}_{[0,1/T]}$ and any $f \in \mathrm{PW}_\Sigma$ can be recovered by sampling at the Nyquist rate of T samples per unit time. Fix a large integer L. For $f \in \mathrm{PW}_\Sigma$ denote $\mathbf{s} = \mathbf{s}(f)$ the sequence with n-th coordinate $s_n = \tilde{f}(nT) = f(-nT)$. For $k = \{0, 1, \ldots, L-1\}$ we define the k-th *sample coset* $\mathbf{s}_k$ of $\mathbf{s}$ to be the sequence with n-th term $s_{kn} = s_n$ if $n = k + mL$ for some $m \in \mathbb{Z}$ and $s_{kn} = 0$ otherwise. In engineering parlance, $\mathbf{s}_k = \sigma^k U_L D_L \sigma^{-k}\mathbf{s}$ where σ is the shift operator $(\sigma\mathbf{s})_n = s_{n-1}$ and D_L and U_L are the respective operators of *downsampling* and *upsampling* by the factor L. To each $\mathbf{s}_k$ one associates its $1/LT$-periodic reversed Fourier series

$$\mathbf{S}_k(\xi) = \sum_n s_{kn}\mathrm{e}^{2\pi i nT\xi} = \mathrm{e}^{2\pi i kT\xi}\sum_{m\in\mathbb{Z}} \widetilde{f}((k+mL)T)\mathrm{e}^{2\pi i mLT\xi} = \ldots$$
$$= \frac{1}{LT}\sum_{\ell=0}^{L-1} \widehat{f}\Big(\xi + \frac{\ell}{LT}\Big)\mathrm{e}^{-2\pi i k\ell/L},\ \xi \in \Big[0, \frac{1}{LT}\Big). \tag{24}$$

The last equality is seen by multiplying both sides by $\mathrm{e}^{-2\pi i kT\xi}$. Then its left hand side becomes the Fourier series of $\sum_{\ell=0}^{L-1} g(\xi + \ell/(LT))$ where $g(\xi) = \frac{1}{LT}\widehat{f}(\xi)\mathrm{e}^{-2\pi i kT\xi}$.

One would like to be able to reconstruct f from P out of L of its sample cosets $\mathbf{s}_k$ where, ideally, $P/L \approx T|\Sigma|$. Then, on average, approximately $|\Sigma|$ samples per unit time are needed to reconstruct f. In what follows we will outline Venkataramani and Bresler's methods for the case of reconstruction from the *first* P out of L cosets. The corresponding sampling method is often referred to as *bunched sampling*. The method for choosing P is called *spectral slicing*.

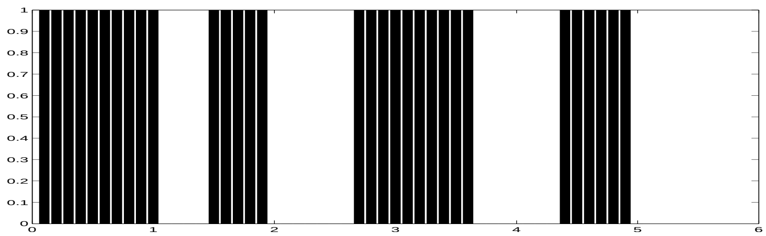

Fig. 5 Multiband spectrum $\Sigma \subset [0, 5]$

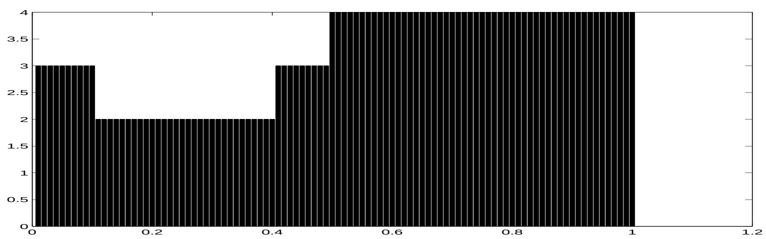

Fig. 6 Pullback of multispectrum, $L = 5$, $T = 1/5$, $1/(LT) = 1$

Spectral slicing. To subdivide the spectrum in an efficient way one starts by denoting the moduli of the endpoints of $\Sigma \subset [0, 1/T]$ modulo $\frac{1}{LT}$ as $0 = \gamma_0 < \gamma_1 < \cdots < \gamma_M$ where $M \le 2N$ since multiple endpoints can have the same modulus. For each $m = 1, \dots, M$ set $\Gamma_m = [\gamma_{m-1}, \gamma_m]$. For each $\ell \in \{0, \dots, L-1\}$, $\Gamma_m + \frac{\ell}{(LT)}$ is either contained in one of the intervals in Σ or is disjoint from Σ. Define index sets $\mathcal{I}_m = \left\{\ell : \Gamma_m + \frac{\ell}{(LT)} \subset \Sigma\right\}$ and $\mathcal{J}_m = \left\{\ell : \Gamma_m + \frac{\ell}{(LT)} \cap \Sigma = \emptyset\right\}$. Spectral slicing is illustrated in Figs. 5 and 6.

Interpolating functions. Let $P = \max_{m=1,\dots,M} |\mathcal{I}_m|$. For each $m = 1, \dots, M$ define a partial Fourier matrix F_m by retaining entries in the first P rows and the columns of the $L \times L$ DFT matrix corresponding to $\ell \in \mathcal{I}_m$, and setting all other entries equal to zero (here one labels the columns from 0 to $L-1$). For the purpose of constructing interpolating functions it is not strictly necessary to use the first P rows – any fixed choice of P rows such that all of the resulting F_m have full rank $|\mathcal{I}_m|$ is sufficient. Denote by F_m^{-1} a fixed partial left inverse of F_m (that is, $F_m^{-1} F_m \mathbf{x} = \mathbf{x}$ whenever $x_\ell = 0$, $\ell \notin \mathcal{I}_m$) and let G_m be any left annihilator of F_m, that is, $G_m F_m = 0$.

The first step in defining appropriate interpolating functions is to extract the spectral components. For each $\ell \in \{0, \dots, L-1\}$ set

$$\widehat{f}_\ell(\xi) = \widehat{f}\Big(\xi + \frac{\ell}{LT}\Big) \mathbb{1}_{\Gamma_m}, \qquad (\ell \in \mathcal{I}_m)$$

(one can take $\widehat{f}_\ell = 0$ if $\ell \notin \cup \mathcal{I}_m$). Since (24) says that $\mathbf{S}_k(\xi)$ is (up to normalization) the $L \times L$ DFT of the sequence $\{\widehat{f}(\xi + \ell/LT)\}$, one can recover $\widehat{f}_\ell$ ($\ell \in \mathcal{I}_m$) from the components $\mathbf{S}_k$, $k = 0, \dots, P-1$ via

$$\widehat{f}_\ell = LT \sum_{k=0}^{P-1} (F_m)^{-1}_{\ell k} \mathbf{S}_k \qquad \text{on } \Gamma_m.$$

Then expressing $f = \sum f_\ell$ one obtains for $\xi \in [0, 1/T]$,

$$\widehat{f}(\xi) = \sum_{m=1}^{M} \sum_{\ell \in \mathcal{I}_m} \Big(LT \sum_{k=0}^{P-1} (F_m)^{-1}_{\ell k} \mathbf{S}_k(\xi) \mathbb{1}_{\Gamma_m}\Big(\xi - \frac{\ell}{LT}\Big)\Big).$$

Taking inverse Fourier transforms then yields

$$f(t) = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{P-1} f\big((k+nL)T\big)\varphi_k(t+nLT) \quad \text{where} \tag{25}$$

$$\varphi_k(t) = LT \sum_{m=1}^{M} \sum_{\ell \in \mathcal{I}_m} (F_m)^{-1}_{\ell k} (\mathbb{1}_{\Gamma_m})^\vee (t+kT)\, \mathrm{e}^{2\pi i \ell (t+kT)/(LT)} \tag{26}$$

which expresses f as a sum over the first P of its L sample cosets. This choice corresponds to taking the left annihilator G_m above to be the zero matrix. One can define more general interpolating functions by taking

$$\varphi_k(t-kT) = LT \sum_{m=1}^{N} \Big(\sum_{\ell \in \mathcal{I}_m} (F_m)^{-1}_{\ell k} \mathrm{e}^{2\pi i \ell t/(LT)} + \sum_{\ell \in \mathcal{J}_m} (G_m)_{\ell k} \mathrm{e}^{2\pi i \ell t/(LT)} \Big) \mathbb{1}^\vee_{\Gamma_m}(t).$$

The matrices G_m might serve to construct interpolating functions with better localization properties. A detailed analysis of the aliasing errors that arise from these techniques is also provided in [40].

Issues with spectral slicing. The parameter $L \in \mathbb{N}$ is called the *slicing parameter*. Complexity of the coset sampling depends on a judicious choice of L as the following examples illustrate.

Example. Consider $\Sigma = [0, 1/p] \cup [1 - 1/q, 1]$. Since $\Sigma \subset [0, 1]$ the Nyquist rate is $T = 1$. The *Landau rate* on the other hand is $|\Sigma| = (p+q)/(pq)$ samples per unit time. This rate can be achieved with $L = pq$.

Example. Let $\Sigma = [0, 1/2] \cup [2/3 - \epsilon, 1 - \epsilon]$. The Nyquist rate is $1 - \epsilon \approx 1$ for small ϵ and the Landau rate is $5/6$. Taking $L = 6$ will yield $P = 6$ so there is no improvement in the sampling rate. If $\epsilon = p/q$ then $L = \mathrm{lcm}(6, q)$ will yield a sampling rate of $5/6$. In this case it is questionable whether the improvement in sampling rate merits the increase in computational complexity. In summary, the distribution of Σ plays a prominant role in the tradeoff between sampling efficiency and computational complexity.

7.2 Layered lattices (Behmard and Faridani)

Periodic nonuniform sampling is one means of reproducing multiband signals from a fixed set of interpolating functions. An alternative approach is to identify a lattice that *fits* each frequency support interval and then peel off components of $f \in \mathrm{PW}_\Sigma$ by resampling remainder terms not accounted for by lattice samples from previous layers. As before, let $\Sigma = \cup [a_k, b_k]$. In algorithmic terms, first one samples on $\mathbb{Z}/(b_1 - a_1)$, then subtracts a suitable interpolant, and iterates on the remainder, sampling along $\mathbb{Z}/(b_2 - a_2)$ etcetera. Care must be taken when the lattices intersect – when the interval lengths are rational multiples of one another.

As an example, suppose that $\Sigma = I_1 \cup I_2$ with $|I_1| = 1/2$ and $|I_2| = 1/3$. To account for I_1 one should sample along $2\mathbb{Z}$ and for I_2 one should sample along $3\mathbb{Z}$. The problem is that there is no way to disentangle information associated with I_1 and I_2 contained in the samples along the intersection $6\mathbb{Z}$. This suggests that, instead, one might sample along $2\mathbb{Z} \cup (3\mathbb{Z} + \alpha)$ for $\alpha \neq 0$.

This is the sort of approach that underlies the work of Behmard and Faridani [3]. Their results are phrased in the language of a locally compact abelian group G, which is a natural context for their approach. But the essence of their results is already captured in the case $G = \mathbb{R}$, as reviewed here. The following is a corollary of Theorem 2 of [3].

Theorem 8 *Let $I_1 \subset I_2$ and let $\Sigma = I_2 \cup (\eta + I_1)$ express Σ as a disjoint union of two intervals. Fix x_1 and x_2 such that for all $n \in \mathbb{Z}$, $\eta(x_1 - x_2 + \frac{n}{|I_1|}) \notin \mathbb{Z}$. Set $\varphi_i = (\mathbb{1}_{I_i})^\vee$, $i = 1, 2$, and let*

$$S_2 f(x) = \sum_{m \in \mathbb{Z}} f\Big(x_2 + \frac{m}{|I_2|}\Big)\varphi_2\Big(x - x_2 - \frac{m}{|I_2|}\Big).$$

Then one can write $f = S_2 + g$ where

$$g(x) = \big(1 - e^{2\pi i (x - x_2)\eta}\big) \sum_{n \in \mathbb{Z}} g\Big(x_1 + \frac{n}{|I_1|}\Big) \frac{\varphi_1\Big(x - x_1 - \frac{n}{|I_1|}\Big)}{1 - e^{2\pi i (x_1 - x_2 + \frac{n}{|I_1|})\eta}}.$$

Example. Let $I_1 = [0, 1/3] \subset [0, 1/2] = I_2$ and set $\eta = 2/3 - \epsilon$ so $\Sigma = [0, 1/2] \cup [2/3 - \epsilon, 1 - \epsilon]$. Let $x_1 = 1/2$ and $x_2 = 0$. Then we can write $S_2 f(x) = \sum_{m \in \mathbb{Z}} f(2m)\varphi_{[0,1/2]}(x - 2k)$ where $\widehat{\varphi_{[a,b]}} = \mathbb{1}^\vee_{[a,b]}$, and then

$$g(x) = \big(1 - e^{2\pi i x(2/3 - \epsilon)}\big) \sum_{n \in \mathbb{Z}} \frac{g\big(3n + \frac{1}{2}\big)}{1 - c_\epsilon e^{-6\pi i n \epsilon}} \varphi_{[2/3 - \epsilon, 1 - \epsilon]}\Big(x - \frac{1}{2} - 3n\Big)$$

where $c_\epsilon = e^{\pi i (2/3 - \epsilon)}$. Expressing $f = g + S_2 f$ shows that f can be recovered from its samples along $2\mathbb{Z} \cup (\frac{1}{2} + 3\mathbb{Z})$, therefore at an average rate of 5 out of every 6 samples. The average sampling rate is thus optimal as one also obtains with the Venkataramani and Bresler methods with large L, but the sampling set here does not correspond to cosets as defined by Venkataramani and Bresler. Theorem 8 can be extended to unions of more than two lattices. Then one has to reiterate the process that defines g as a remainder.

The Venkataramani-Bresler and Behmard-Faridani theorems are examples of how to obtain sampling sets $\{x_n\}$ and interpolating functions ψ_n that give rise to (16) as mentioned also in Problem 9. The distinct sampling methods will also lead to distinct formulations of sampling projections onto spaces of approximately time and frequency localized functions.

A Outline of Proof of Proposition 3

A.1 Relational calculus on $\mathbb{Z}_N^{2n}$.

Partitions of $\{1,\dots,N\}$. A partition P of $A=\{1,\dots,N\}$ is a collection $U_1,\dots U_k$ of subsets of A that are pairwise disjoint ($U_i\cap U_j=\emptyset$) and such that each element of A is contained in one of the U_k ($\cup U_k=A$). A partition can also be thought of as a decomposition of an equivalence relation into its equivalence classes and this is the approach taken by Candès, Romberg and Tao [9]. In the following outline of their arguments we translate their *relational calculus* into the language of *partitions*. When $P=\{U_1,\dots,U_k\}$ we will write $U_j\in P$ and we will denote by $|P|$ the number k of nonempty subsets (classes) into which P partitions A. A partition $Q=\{U_1',\dots,U_\ell'\}$ is called a *refinement* of $P=\{U_1,\dots,U_k\}$ – expressed by $P\le Q$ – if for each $U_i'\in Q$ there is a $U_j\in P$ such that $U_i'\subset U_j$. In other words, Q repartitions each of the partition elements of P. Finally, we will write $\mathcal{P}=\mathcal{P}(A)$ for the collection of all partitions of A partially ordered by refinement.

Stirling numbers. Partitions of finite sets satisfy certain combinatorial properties. For example, the *Stirling numbers of the second kind* $\mathrm{St}\,(n,k)$ enumerate the number of distinct ways in which a set of n elements can be partitioned into k pairwise disjoint subsets. These numbers satisfy

$$\mathrm{St}\,(n+1,k)=\mathrm{St}\,(n,k-1)+k\,\mathrm{St}\,(n,k)$$

since, if a is a distinguished element of A (where $|A|=n+1$) then $\mathrm{St}\,(n,k-1)$ counts all of those k-partitions containing the singleton a as one of its elements while $k\,\mathrm{St}\,(n,k)$ counts the k different ways that a can be added to one of the elements of each of the k-partitions of $A\setminus\{a\}$.

Function partitions. If A and B are finite sets then any function $f:A\to B$ partitions A by setting $U_j=f^{-1}(\beta_j)$ and $P_f=\{U_1,\dots,U_k\}$. Here $\beta_1,\dots\beta_k$ refers to an enumeration of the range of f. If σ is any permutation of the elements of B then $g=\sigma\circ f$ gives rise to the same partition $P=P_f=P_g$ where, now, $U_j=g^{-1}(\sigma(\beta_j))$ so the enumeration of the partition sets might be different but the partition elements will be exactly the same. In fact, if $P_f=P_g$ then g has the form $\sigma\circ f$ for some permutation σ of the range. A lattice structure on the functions mapping A to B can be defined by $f\ll g$ if $P_f\le P_g$ (i.e., P_g is a refinement of P_f).

Expected values. Suppose it is stipulated that f has a *random* frequency support in which each frequency is a Bernoulli random variable. In what follows we consider N point DFTs and denote by Ω a subset of $\mathbb{Z}_N$ that is randomly generated with each ω_j, $j=0,\dots,N-1$ belonging to Ω with fixed probability $\tau=\mathbf{E}(|\Omega|)/N$ for a desired expected value of $|\Omega|$. Thus Ω is defined in terms of N Bernoulli random variables $I(\omega_j)$ taking value zero if $\omega_j\notin\Omega$ and one if $\omega_j\in\Omega$, each with probability τ. We will sometimes write $I(\omega_j)=\mathbb{1}_\Omega(\omega_j)$ where the latter indicates explicitly whether $\omega_j\in\Omega$.

A.2 Traces of powers Powers of $H_{S\Omega}$

Recall that $A_{S\Omega}=D_S\mathcal{F}_N^{-1}D_\Omega\mathcal{F}_N$ is the finite analogue of Q_SP_Ω. One has $A_{S\Omega}A_{S\Omega}^*=D_S\mathcal{F}_N^{-1}D_\Omega\mathcal{F}_ND_S$ and the auxiliary matrix $H=H_{S\Omega}$ where

$$H(s,t)=\begin{cases}0, & s=t\\ c(s-t) & s\ne t,\ c(u)=\sum_{\omega\in\Omega}\mathrm{e}^{\frac{2\pi i}{N}\omega u}\end{cases}$$

satisfies $H_{S\Omega}=NA_{S\Omega}A_{S\Omega}^*-|\Omega|I$.

The function on $\mathbb{Z}_N$ obtained by taking the trace of a power of $H=H_{S\Omega}$ is far from being an arbitrary function and one takes advantage of this in estimating the likelihood of H having a large norm.

A diagonal element of the $2n$-th power of H can be written

$$H^{2n}(t_1,t_1)=\sum_{t_2,\ldots,t_{2n}:t_j\neq t_{j+1}}\prod_j c(t_j-t_{j+1})$$

where $t_{2n+1}\equiv t_1$. The expected value of the trace of H^{2n} is that of the sum over these diagonal elements, namely

$$\begin{aligned}\mathbf{E}(\mathrm{tr}(H^{2n})) &= \sum_{t_2,\ldots,t_{2n}:t_j\neq t_{j+1}}\mathbf{E}\Bigg[\sum_{\omega_1,\ldots,\omega_{2n}\in\Omega}\mathrm{e}^{\frac{2\pi i}{N}\sum_{j=1}^{2n}\omega_j(t_j-t_{j+1})}\Bigg]\\ &= \sum_{t_2,\ldots,t_{2n}:t_j\neq t_{j+1}}\ \sum_{0\le\omega_1,\ldots,\omega_{2n}<N}\mathrm{e}^{\frac{2\pi i}{N}\sum_{j=1}^{2n}\omega_j(t_j-t_{j+1})}\mathbf{E}\Bigg[\prod_{j=1}^{2n}I(\omega_j)\Bigg]\end{aligned}$$

where one has applied the linearity of expectation together with the definition $I(\omega)=1$ if $\omega\in\Omega$, where $I(\omega)$ is regarded as a $\{0,1\}$-Bernoulli random variable with expected value $\tau\in(0,1)$.

Expectation conditioned on a partition. Any vector $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_{2n})\in\mathbb{Z}_N^{2n}$ can be regarded as a function from $\{1,2,\ldots,2n\}$ to $\mathbb{Z}_N$. As such, $\boldsymbol{\omega}$ induces a partition $P_{\boldsymbol{\omega}}$ of $\{1,\ldots,2n\}$ whose partition elements are the inverse images of the frequency elements $\omega_j\in\mathbb{Z}_N$. Once $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_{2n})$ is fixed the values of $I(\omega_j)$ and $I(\omega_k)$ are equal if $\omega_j=\omega_k$ so one can think of $\boldsymbol{\omega}$ and its partition $P=P_{\boldsymbol{\omega}}$ as inducing a *conditional expectation* on $\{1,2,\ldots,2n\}$ by forcing $I(\omega_j)=I(\omega_k)$ whenever j,k belong to the same partition set. Then

$$\mathbf{E}\Bigg[\prod_{j=1}^{2n}I(\omega_j)\Bigg]=\tau^{|P|}$$

whenever $P_{\boldsymbol{\omega}}=P$. Thus coarser partitions yield higher expected values.

Given a fixed partition P, denote by $W(P)$ those $\boldsymbol{\omega}\in\mathbb{Z}_N^{2n}$ such that $P_{\boldsymbol{\omega}}=P$. Then

$$\mathbf{E}(\mathrm{tr}(H^{2n}))=\sum_{t_2,\ldots,t_{2n}:t_j\neq t_{j+1}}\ \sum_{P\in\mathcal{P}}\tau^{|P|}\sum_{\boldsymbol{\omega}\in W(P)}\mathrm{e}^{\frac{2\pi i}{N}\sum_{j=1}^{2n}\omega_j(t_j-t_{j+1})}$$

A.3 Combinatorics for exponential sums

Candès, Romberg and Tao introduced at this stage certain *inclusion-exclusion formulae* that allow simplification of the exponential sums inside the sum over the partitions. In what follows, for $V\subset A$, $P(V)$ denotes the partition of V induced by the partition P of F.

Lemma 3 (Inclusion-exclusion) *Let A and B be finite, nonempty subsets. Then for any $P\in\mathcal{P}(A)$ and any $f:B^{|A|}\to\mathbb{C}$,*

$$\sum_{\boldsymbol{\omega}\in W(P)}f(\boldsymbol{\omega})=\sum_{Q\le P}(-1)^{|P|-|Q|}\Big(\prod_{U\in Q}(|P(U)|-1)!\Big)\sum_{R\le Q}\ \sum_{\boldsymbol{\omega}\in W(R)}f(\boldsymbol{\omega}).$$

By splitting A into its partition elements with respect to a fixed Q one has

$$\sum_{P:Q\le P}\tau^{|P|}(-1)^{|P|-|Q|}\Big(\prod_{U\in Q}(|P(U)|-1)!\Big)=\prod_{U\in Q}\sum_{k=1}^{|U|}\mathrm{St}\,(|U|,k)\,\tau^k\,(-1)^{|U|-k}(k-1)!$$

where, as before, $\mathrm{St}\,(n,k)$ is the Stirling number. The inner sum depends only on $|U|$ and τ so, defining $F(n,\tau)=\sum_{k=1}^{n}\mathrm{St}\,(n,k)\,\tau^k\,(-1)^{n-k}(k-1)!$ we see that

$$\sum_{P:Q\le P}\tau^{|P|}(-1)^{|P|-Q|}\Big(\prod_{U\in Q}(|P(U)|-1)!\Big)=\prod_{U\in Q}F(|U|,\tau).$$

In summary,

Lemma 4

$$\sum_{P\in\mathcal{P}} \tau^{|P|} \sum_{\boldsymbol{\omega}\in W(P)} f(\boldsymbol{\omega}) = \sum_{Q\in\mathcal{P}} \left[\sum_{R\leq Q} \sum_{\boldsymbol{\omega}\in W(R)} f(\boldsymbol{\omega}) \right] \prod_{U\in Q} F(|U|,\tau).$$

Specializing to $f(\boldsymbol{\omega}) = \mathrm{e}^{\frac{2\pi i}{N}\sum_{1\leq j\leq 2n}\omega_j(t_j-t_{j+1})}$ yields the following formula for the expected trace of H^{2n}:

$$\mathbf{E}(\mathrm{tr}(H^{2n})) = \sum_{P\in\mathcal{P}} \sum_{t_2,\ldots,t_{2n}:t_j\neq t_{j+1}} \sum_{R\leq P} \sum_{\boldsymbol{\omega}\in W(R)} \mathrm{e}^{\frac{2\pi i}{N}\sum_{j=1}^{2n}\omega_j(t_j-t_{j+1})} \prod_{U\in P} F(|U|,\tau).$$

Fix P and $U\in P$ and set $t_U = \sum_{j\in U}(t_j - t_{j+1})$. For $\boldsymbol{\omega}\in W(R)$ ($R\leq P$) the value of ω_j then depends only on the partition set containing j so we write $\omega_j = \omega_U$; however, as R ranges over $R\leq P$ and $\boldsymbol{\omega}$ ranges over $W(R)$ the set of $\{\boldsymbol{\omega}_U = (\omega_{U_1},\ldots,\omega_{U_{|P|}})\}$ ranges precisely over $\mathbb{Z}_N^{|P|}$. In light of this we have

$$\sum_{R\leq P} \sum_{\boldsymbol{\omega}\in W(R)} \mathrm{e}^{\frac{2\pi i}{N}\sum_{U\in P}\omega_U t_U} = \sum_{\boldsymbol{\omega}_U\in\mathbb{Z}_N^{|P|}} \mathrm{e}^{\frac{2\pi i}{N}\sum \omega_U t_U} = \begin{cases} N^{|P|}, & t_U = 0 \text{ all } U \\ 0, & t_U\neq 0 \text{ some } U \end{cases}.$$

Substituting this back into the expected trace formula above leads to the following.

Lemma 5

$$\mathbf{E}(\mathrm{tr}(H^{2n})) = \sum_{P\in\mathcal{P}} \sum_{t_2,\ldots,t_{2n}:t_j\neq t_{j+1} \text{ and } \forall U\in P, t_U=0} N^{|P|} \prod_{U\in P} F(|U|,\tau).$$

This represents a rather dramatic simplification of the expression for $\mathbf{E}(\mathrm{tr}(H^{2n}))$. However, it remains to estimate the number of terms for which t_U vanishes for all $U\in P$ and then to estimate $\prod_{U\in P} F(|U|,\tau)$. The first point is addressed as follows.

Lemma 6 *For any $P\in\mathcal{P}$,*

$$\#\{t\in S^{2n} : t_U = 0 \text{ for all } U\in P\} \leq |S|^{2n-|P|+1}.$$

Any further estimates on $\mathbf{E}(\mathrm{tr}(H^2 n))$ reduce to estimating $F(n,\tau)$ and to counting the number of partitions of $\{1,\ldots,2n\}$ whose partition set sizes have a given distribution. The starting point is the following formula for $F(n,\tau)$ which follows from the recursion relation for the Stirling numbers:

$$F(n,\tau) = \sum_{k=1}^{n} (-1)^{n+k} \frac{\tau^k k^{n-1}}{(1-\tau)^k}.$$

With this formula one can obtain bounds on $F(n,\tau)$ once the size of τ is fixed. For example

Lemma 7 *Let $n\geq 1$ and $0\leq\tau<1/2$ and set $\tau^* = \frac{\tau}{1-\tau}$. Define*

$$G(n+1) \equiv \begin{cases} \tau^* & \text{if } \tau\leq \mathrm{e}^{-n} \\ \mathrm{e}^{n(\log n - \log\log\frac{1}{\tau^*} - 1)} & \text{if } \tau > \mathrm{e}^{-n} \end{cases}$$

Then $|F(n,\tau)|\leq G(n)$.

Applying Lemmas 5 and 6 one concludes that

$$\mathbf{E}(\mathrm{tr}(H^{2n})) \leq \sum_{k=1}^{n} N^k |S|^{2n-k+1} \sum_{P:\,|P|=k} \prod_{U\in P} G(|U|). \tag{27}$$

Convexity of $\log G(n)$ together with more partition combinatorics and induction lead to the estimate

$$\sum_{P:\,|P|=k} \prod_{U\in P} G(|U|) \leq \frac{(n-1)!}{(k-1)!}\, 2^{n-2k+1}\, G(2)^k$$

and summing over k in (27) results in the estimate

$$\mathbf{E}(\mathrm{tr}(H^{2n})) \leq n\frac{2n!}{n!}G(2)^n|S|^{n+1}N^n.$$

Applying the classical Stirling approximation then yields

$$\mathbf{E}(\mathrm{tr}(H^{2n})) \leq n2^{2n+1}n^n\mathrm{e}^{-n}G(2)^n|S|^{n+1}N^n.$$

This completes our outline of the proof of Proposition 3.

Acknowledgements The authors would like to thank Q. Sun for suggesting that we prepare an outline of connections open problems regarding sampling and time-frequency localization of multiband signals. They would also like to thank Chuck Creusere, Chris Brislawn and Jeff Hogan for several pertinent discussions.

References

1. William Arveson. *A short course on spectral theory*, volume 209 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 2002.
2. Louis Auslander and F. Alberto Grünbaum. The Fourier transform and the discrete Fourier transform. *Inverse Problems*, 5(2):149–164, 1989.
3. Hamid Behmard and Adel Faridani. Sampling of bandlimited functions on unions of shifted lattices. *J. Fourier Anal. Appl.*, 8(1):43–58, 2002.
4. Hamid Behmard, Adel Faridani, and David Walnut. Construction of sampling theorems for unions of shifted lattices. *Sampl. Theory Signal Image Process.*, 5(3):297–319, 2006.
5. M. Benedicks. On the Fourier transform of functions supported on sets of finite Lebesgue measure. *J. Math. Anal. Appl.*, 106:180–183, 1985.
6. G. Beylkin and L. Monzón. On generalized Gaussian quadratures for exponentials and their applications. *Appl. Comput. Harmon. Anal.*, 12:332–373, 2002.
7. Emmanuel Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.
8. Emmanuel J. Candès and Justin Romberg. Errata for: "Quantitative robust uncertainty principles and optimally sparse decompositions" [Found. Comput. Math. **6** (2006), no. 2, 227–254; mr2228740]. *Found. Comput. Math.*, 7(4):529–531, 2007.
9. Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
10. D. Donoho and P. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49:906–931, 1989.
11. D. Donoho and P. Stark. A note on rearrangements and spectral concentration. *IEEE Trans. Inform. Theory*, 39:257–260, 1993.
12. Yonina C. Eldar and Alan V. Oppenheim. Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples. *IEEE Trans. Signal Process.*, 48(10):2864–2875, 2000.
13. C. Herley and P.-W. Wong. Minimum rate sampling and reconstruction of signals with arbitrary frequency support. *IEEE Trans. Inform. Theory*, 45:1555–1564, 1999.
14. J. R. Higgins. Five short stories about the cardinal series. *Bull. Amer. Math. Soc. (N.S.)*, 12(1):45–89, 1985.
15. J. R. Higgins. Sampling for multi-band functions. In *Mathematical analysis, wavelets, and signal processing (Cairo, 1994)*, volume 190 of *Contemp. Math.*, pages 165–170. Amer. Math. Soc., Providence, RI, 1995.
16. J. R. Higgins. Historical origins of interpolation and sampling, up to 1841. *Sampl. Theory Signal Image Process.*, 2(2):117–128, 2003.
17. Jeffrey A. Hogan and Joseph D. Lakey. Sampling and time-frequency localization of bandlimited and multiband signals. In *Wavelets and Frames: A Celebration of the Mathematical Work of Lawrence Baggett*, Appl. Numer. Harmon. Anal. Birkhäuser Boston, Boston, MA. to appear.

18. Jeffrey A. Hogan and Joseph D. Lakey. *Time-frequency and time-scale methods*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2005. Adaptive decompositions, uncertainty principles, and sampling.
19. Scott Izu. *Sampling and Time-Frequency Localization in Paley-Wiener Spaces*. PhD thesis, New Mexico State University. in preparation.
20. M. Kac, W. L. Murdock, and G. Szegö. On the eigenvalues of certain Hermitian forms. *J. Rational Mech. Anal.*, 2:767–800, 1953.
21. Kedar Khare and Nicholas George. Sampling theory approach to prolate spheroidal wave-functions. *J. Phys. A*, 36(39):10011–10021, 2003.
22. H. J. Landau. The eigenvalue behavior of certain convolution equations. *Trans. Amer. Math. Soc.*, 115:242–256, 1965.
23. H. J. Landau. On Szegő's eigenvalue distribution theorem and non-Hermitian kernels. *J. Analyse Math.*, 28:335–357, 1975.
24. H. J. Landau and H. Widom. Eigenvalue distribution of time and frequency limiting. *J. Math. Anal. Appl.*, 77(2):469–481, 1980.
25. H.J. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.*, 117:37–52, 1967.
26. H.J. Landau. An overview of time and frequency limiting. In J. F. Price, editor, *Fourier Techniques and Applications*, pages 201–220. Plenum Press, New York, 1985.
27. H.J. Landau. On the density of phase-space expansions. *IEEE Trans. Inform. Theory*, 39:1152–1156, 1993.
28. H.J. Landau and H.O. Pollak. Prolate spheroidal wavefunctions, Fourier analysis and uncertainty. II. *Bell Syst. Tech. J.*, 40:65–84, 1961.
29. H.J. Landau and H.O. Pollak. Prolate spheroidal wavefunctions, Fourier analysis and uncertainty. III. The dimension of the space of essentially time- and band-limited signals. *Bell Syst. Tech. J.*, 41:1295–1336, 1962.
30. Josef Meixner and Friedrich Wilhelm Schäfke. *Mathieusche Funktionen und Sphäroidfunktionen mit Anwendungen auf physikalische und technische Probleme*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete, Band LXXI. Springer-Verlag, Berlin, 1954.
31. F. L. Nazarov. Local estimates for exponential polynomials and their applications to inequalities of the uncertainty principle type. *Algebra i Analiz*, 5(4):3–66, 1993.
32. D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty IV. Extensions to many dimensions; generalized prolate spheroidal wave functions. *Bell Syst. Tech. J.*, 43:3009–3057, 1964.
33. D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty V. The discrete case. *Bell Syst. Tech. J.*, 57:1371–1430, 1964.
34. D. Slepian. On bandwidth. *Proc. IEEE*, 64:292–300, 1976.
35. D. Slepian and H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. I. *Bell Systems Tech. J.*, 40:43–64, 1961.
36. David Slepian. Some asymptotic expansions for prolate spheroidal wave functions. *J. Math. and Phys.*, 44:99–140, 1965.
37. J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbató. *Spheroidal wave functions, including tables of separation constants and coefficients*. Technology Press of M. I. T. and John Wiley & Sons, Inc., New York, 1956.
38. Terence Tao. An uncertainty principle for cyclic groups of prime order. *Math. Res. Lett.*, 12(1):121–127, 2005.
39. C. Thiele. Time–frequency analysis in the discrete phase plane. In *Topics in Analysis and its Applications*, pages 99–152. World Scientific, River Edge, NJ, 2000.
40. Raman Venkataramani and Yoram Bresler. Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals. *IEEE Trans. Inform. Theory*, 46(6):2173–2183, 2000.
41. Gilbert G. Walter and Xiaoping A. Shen. Sampling with prolate spheroidal wave functions. *Sampl. Theory Signal Image Process.*, 2(1):25–52, 2003.
42. H. Xiao, V. Rokhlin, and N. Yarvin. Prolate spheroidal wavefunctions, quadrature and interpolation. *Inverse Problems*, 17:805–838, 2001.
43. Ahmed I. Zayed. *Advances in Shannon's sampling theory*. CRC Press, Boca Raton, FL, 1993.