RANDOM SAMPLING OF BAND-LIMITED FUNCTIONS

RICHARD F. BASS AND KARLHEINZ GRÖCHENIG

Abstract. We consider the problem of random sampling for band-limited functions. When can a band-limited function f be recovered from randomly chosen samples $f(x_j), j \in \mathbb{N}$? We estimate the probability that a sampling inequality of the form

$$A\|f\|_2^2 \le \sum_{j\in\mathbb{N}} |f(x_j)|^2 \le B\|f\|_2^2$$

hold uniformly all functions $f \in L^2(\mathbb{R}^d)$ with supp $\hat{f} \subseteq [-1/2, 1/2]^d$ or some subset of band-limited functions.

In contrast to discrete models, the space of band-limited functions is infinite-dimensional and its functions "live" on the unbounded set $\mathbb{R}^d$. This fact raises new problems and leads to both negative and positive results.

(a) With probability one, the sampling inequality fails for any reasonable definition of a random set on $\mathbb{R}^d$, e.g., for spatial Poisson processes or uniform distribution over disjoint cubes.

(b) With overwhelming probability, the sampling inequality holds for certain compact subsets of the space of band-limited functions and for sufficiently large sampling size.

1. Introduction

The sampling problem asks for the reconstruction or approximation of a function f from its sampled values $\{f(x_j) : j \in J\}$ on some set $\mathcal{X} = \{x_j\} \subseteq \mathbb{R}^d$. In other words, one wants to recover f from given samples $f(x_j)$. This is a many-faceted problem and spreads over many areas of mathematics, engineering, and data processing.

We will impose the standard hypothesis that f is band-limited. In signal processing this is a realistic assumption, because it amounts to assuming a maximum frequency. The assumption is also relevant in complex analysis because a band-limited function is just the restriction of an entire function of exponential growth from $\mathbb{C}^d$ to $\mathbb{R}^d$. The space of band-limited functions is defined to be

$$\mathcal{B} = \{f \in L^2(\mathbb{R}^d) : \text{supp}\, \hat{f} \subseteq [-1/2, 1/2]^d\},$$

where we have normalized the spectrum to be the unit cube and the Fourier transform is normalized as $\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i x\cdot\xi}\, dx$.

2000 *Mathematics Subject Classification.* 94A12,94A20,15A12,15A52,42B99,42A15,42A61, 60G50,60G99.

Key words and phrases. Band-limited function, random sampling, condition number, Poisson process, prolate spheroidal functions.

R.B. was partially supported by NSF Grant DMS0601783.

K.G. was supported by the Marie-Curie Excellence Grant MEXT-CT-2004-517154.

The principal goal is to establish a *sampling inequality* of the form

$$A\|f\|_2^2 \leq \sum_j |f(x_j)|^2 \leq B\|f\|_2^2 \qquad \forall f \in \mathcal{B}\,. \tag{1}$$

A set $\{x_j : j \in J\} \subseteq \mathbb{R}^d$ satisfying the sampling inequality (1) is called a set of stable sampling or simply a set of sampling [17]. Once a sampling inequality is established, every $f \in \mathcal{B}$ is uniquely determined by its samples on $\mathcal{X}$ and depends continuously on these samples. For the actual reconstruction of f from its samples one then may use standard algorithms from frame theory [12].

Band-limited functions in dimension $d = 1$ and $d > 1$ differ in a fundamental way because of the nature of their zeros. In dimension $d = 1$ the zeros of an entire function are always discrete, and there is a precise connection between the possible density of zeros and the growth of f [4, 21, 27]. By contrast, in higher dimensions, the zero sets are analytic manifolds, and standard complex variable techniques do no longer apply. As a consequence, almost everything is known about the sampling of band-limited functions in dimension $d = 1$, but only a few results are known in higher dimensions, most notably a strong result of Beurling [3].

The difficulties of the sampling of multivariate functions have motivated us to turn to probabilistic techniques and to study *random sampling*. In this approach the sampling set $\mathcal{X}$ is a sequence of random variables $x_j = x_j(\omega)$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $\mathbb{R}^d$. The sampling inequality (1) defines an event on Ω, and the goal is to estimate the probability that a random set is a set of sampling.

This point of view has worked successfully in our previous work [1] where we have studied the random sampling of multivariate trigonometric polynomials. We were able to show that some popular numerical algorithms [26] work with "overwhelming" probability. In a similar spirit, Candès, Romberg, and Tao [5, 6] have recently investigated sparse trigonometric polynomials and their reconstruction from a few random samples. The more general context of mathematical learning theory has been studied by Cucker, Poggio, Smale, and Zhou [10, 25, 32]. In [32] sampling in general reproducing kernel Hilbert spaces was studied under the assumption of "rich data." This amounts to assuming the validity of a sampling inequality. By contrast, our interest is to establish the probability that this basic assumption holds. The common technical point in these approaches [1, 5, 10] is the estimate of entropy and covering numbers and a metric entropy argument.

The first contributions to random sampling of band-limited functions were perturbation results in dimension $d = 1$. Seip and Ulanovsky [28] investigated random perturbations of regular sampling $\{j + \delta_j : j \in \mathbb{Z}\}$, where δ_j is a sequence of i.i.d. random variables. Chistyakov, Lyubarskii, Pastur [8, 9] studied the more general problem of perturbation of arbitrary Riesz bases of exponentials. These contributions are based on the precise characterization of sampling sets in dimension $d = 1$ [27], and the proofs proceed by estimating the probability that a deterministic condition is satisfied.

For random sampling of band-limited functions of several variables new types of problems arise.

(a) One cannot fall back on deterministic results in higher dimensions, because sampling theory is not nearly as developed as in dimension $d = 1$. In fact, this is the very reason why we aim for purely probabilistic results.

(b) The space of band-limited functions $\mathcal{B}$ is infinite-dimensional — in contrast to trigonometric polynomials of given degree or sparsity. Thus random matrix techniques as used in [13, 23] are not applicable.

(c) The configuration space $\mathbb{R}^d$ is non-compact and unbounded — again in contrast to trigonometric polynomials that "live" on the torus $[0,1]^d$. This raises the question of how to model a sequence of random points in $\mathbb{R}^d$. On a compact set of positive (Lebesgue) measure the natural notion is that of an independent identically distributed (i.i.d.) sequence of points with uniform distribution. On $\mathbb{R}^d$ there are several natural choices. We will consider two such choices: uniform distributions on disjoint cubes and spatial Poisson processes.

We will prove that for these two concepts of "randomly distributed points on $\mathbb{R}^d$" the sampling inequality (1) must fail almost surely (Propositions 2.2 and 2.3). These results come as a surprise to the analyst, but are perhaps more natural for the probabilist. The reasons for the failure of a sampling inequality are either the zeros of entire functions or large holes in the sampling set. In the model of uniform distribution over disjoint cubes, many samples may be near the zeros of a band-limited functions with positive probability. In other words, the lower bound in (1) is small.

In the other model (spatial Poisson process) we show that, with positive probability, there are large holes in the sampling set, which again implies a small lower bound in (1).

To obtain insight into the formulation of positive results, we argue in a practical manner. Realistically one can sample f only on a bounded set; furthermore, every band-limited function vanishes at infinity, thus samples far out do not contribute anything significant to a sampling inequality. We can learn about $f \in \mathcal{B}$ only if the samples are taken in the "essential support" of f, i.e., the set where most of the L^2-norm is localized. Thus we will study the subset

$$\mathcal{B}(R,\delta) = \{f \in \mathcal{B} : \int_{[-R/2,R/2]^d} |f(x)|^2 \, dx \geq (1-\delta)\|f\|_2^2\}$$

of band-limited functions. This subset is compact in $\mathcal{B}$ and thus somewhat resembles a finite-dimensional subspace. Since $f \in \mathcal{B}(R,\delta)$ is small outside the cube $[-R/2, R/2]^d$, it should suffice to sample f on the relevant cube. In this way, we are back to a compact configuration space and an almost finite-dimensional function space. Our main result (Theorem 3.1) is a restricted sampling inequality for the subset $\mathcal{B}(R,\delta)$. The proof is a combination of analytic and probabilistic techniques. On the one hand, we will use detailed properties about the spectrum of time-limiting operators on band-limited functions by Widom [33], on the other hand, the metric entropy method (see e.g., [11]). In order to determine for which sample sizes the probability estimates become effective, we will keep track of the explicit numerical constants.

The paper is organized as follows. In Section 2 we discuss two natural models for random sequences in $\mathbb{R}^d$ and show that, with probability one, they fail to produce sets of stable sampling. In Section 3 we restrict the attention to a subset of band-limited functions and show that on this subset a sampling inequality holds with overwhelming probability. The proof of this result is contained in Section 4. In order to set up the metric entropy method, we discuss the spectrum of time-frequency limiting operators and covering numbers.

2. Negative Results

In the case of multivariate trigonometric polynomials, we showed that if one chose points independently and uniformly distributed over the state space, then one could recover the trigonometric polynomial exactly provided only that one had at least as many sample points as the dimension [1, Thm 3.2].

We first show that this is far from the case for band-limited functions. The difficulty is that the state space is not compact.

We first recall a fundamental necessary condition of Landau for a set of sampling. Let

$$D^-(\mathcal{X}) = \lim_{R\to\infty} \min_{y\in\mathbb{R}^d} \frac{\operatorname{card} \mathcal{X} \cap (y+[0,R]^d)}{R^d} \tag{2}$$

be the (lower) Beurling density of a set $\mathcal{X} \subseteq \mathbb{R}^d$.

Proposition 2.1. *Assume that $\mathcal{X} = \{x_j\}$ is a set of stable sampling for $\mathcal{B}$. Then $\mathcal{X}$ satisfies*
(i) $D^-(\mathcal{X}) \geq 1$,
(ii) the number of samples in any cube of length 1 is bounded, $\max_{y\in\mathbb{R}^d} \operatorname{card} \mathcal{X} \cap (y+[0,1]^d) < \infty$, *and*
(iii) there is $R>0$ such that every cube of side length R contains a sampling point, i.e.

$$\mathcal{X} \cap (x+[0,R]^d) \neq \emptyset \qquad \forall x \in \mathbb{R}^d\,.$$

A sufficient condition is the following: If $d=1$ and $D^-(\mathcal{X}) > 1$, then $\mathcal{X}$ is a set of sampling.

Proof. (i) and (iii) are the results of Landau [17] and have been re-derived in [14] for discrete sampling sets; the general case is an easy extension. For (iii) see also [15].

(ii) is an easy consequence of the finiteness of upper bound B in (1).

The sufficient condition in dimension $d=1$ is usually attributed to Beurling and treated in detail by Seip [27]. □

Loosely speaking, a set of stable sampling must be dense enough and cannot have arbitrarily large "holes".

We now consider random sampling sets. Let our probability space be $(\Omega, \mathcal{F}, \mathbb{P})$ and denote points in Ω by ω. When sampling a function f randomly, we consider its samples $f(x_j)$ on a sequence of random points $x_j = x_j(\omega)$. Clearly, a sequence of random points need not have the sufficient density stated in Proposition 2.1. However, if the process is designed to yield only random sets with $D^-(\mathcal{X}) \geq 1$, one could hope that a generic random set with the necessary density would be a set of stable sampling. This intuition is completely false, as we will show in the next sections.

2.1. **Uniform distribution on large disjoint cubes.** There are various ways in which one could choose points randomly in $\mathbb{R}^d$. As a first model we partition $\mathbb{R}^d$ into disjoint cubes $k+[0,1]^d, k \in \mathbb{Z}^d$, and, in each cube, we choose r points independently and uniformly distributed over $k+[0,1]^d$. Let $\mathcal{X}$ be the collection of sample points; $\mathcal{X}$ is a random set and thus depends on ω. Clearly $D^-(\mathcal{X}) = r$ almost surely, so one may expect that $\mathcal{X}(\omega)$ is a set of stable sampling with high probability.

Our first result says that one cannot obtain a sampling inequality.

Proposition 2.2. *Let $r \geq 1$ be the number of random samples in each cube $k+[0,1]^d$. With probability one the following holds:*
For each $k > 0$ there is a function $f_k \in \mathcal{B}$ such that

$$\sum_{x_i \in \mathcal{X}(\omega)} |f_k(x_i)|^2 \leq \frac{1}{k} \|f_k\|_2^2.$$

The function f_k will necessarily depend on ω.

Consequently, a sampling inequality of the form (1) *is violated almost surely.*

Proof. For notational simplicity we give the proof only in dimension $d = 1$; the case of several variables is treated similarly.

Let

$$g(x) = \frac{\sin(\pi x/2)}{\pi x/2},$$

and let ψ be a nonnegative C^∞ function with support in $[-1/4, 1/4]$ such that $\psi = 1$ on $[-1/8, 1/8]$. Let Ψ be the inverse Fourier transform of ψ and define $F(x) = g(x)\Psi(x)$.

Since ψ and thus Ψ are in the Schwartz class, F is in L^2, decays rapidly, and there exists a constant c_1 such that $|F(x)| \leq c_1/(1+|x|^2)$. The Fourier transform of F is $\widehat{g} * \psi$, so the support of $\widehat{F}$ lies in $[-1/2, 1/2]$, i.e., $F \in \mathcal{B}$. Since F is bounded, by Bernstein's inequality, F' is also bounded, say, by $c_2 := \|F'\|_\infty$.

Choose N a large even integer so that

$$\sum_{|j| \geq N/2} \frac{c_1^2 r}{(1+(|j|-1)^2)^2} < \frac{\|F\|_2^2}{4k}. \tag{3}$$

Choose $\delta > 0$ small so that

$$2c_2 N r \delta < \frac{\|F\|_2^2}{2k}. \tag{4}$$

Let A_j be the event that in the interval $[j, j+1]$ all r points that were chosen randomly lie within $(j, j+\delta)$ if j is even and within $(j+1-\delta, j+1)$ if j is odd. The events A_j are independent and the probability of A_j is δ^r.

Let $B = \bigcap_{j=-N}^{N} A_j$ be the event that the samples in $[-N, N]$ are in a δ-neighborhood of the even integers. By independence, the probability of B is $(\delta^r)^{2N+1}$. If $\omega \in B$, then using (3) and our bound on F,

$$\sum_{x_i \in \mathcal{X}(\omega) \setminus [-N,N]} |F(x_i)|^2 \le \sum_{|j| \ge N/2} \frac{c_1^2 r}{(1 + (|j|-1)^2)^2} < \frac{\|F\|_2^2}{4k}.$$

By construction, $F(2j) = 0$ for $j \in \mathbb{Z}$, and so using the bound on F', we have $|F(x)| \le c_2 \delta$ if $|x - 2j| \le \delta$. Therefore if $\omega \in B$, then

$$\sum_{x_i \in \mathcal{X}(\omega) \cap [-N,N]} |F(x_i)|^2 \le 2c_2 N r \delta < \frac{\|F\|_2^2}{2k}.$$

Combining, if $\omega \in B$, then

$$\sum_{x_i \in \mathcal{X}(\omega)} |F(x_i)|^2 < \frac{\|F\|_2^2}{k}. \tag{5}$$

Now let $C_m = \bigcap_{j=3mN-N}^{3mN+N} A_j$. Clearly the probability of C_m is the same as the probability of B. So $\sum_{m=1}^{\infty} \mathbb{P}(C_m) = \infty$. By independence and the Borel-Cantelli lemma, with probability one, C_m occurs for infinitely many m. If $\omega \in C_m$, let $f_k(x) = F(x - 3mN)$. Clearly, $f_k \in \mathcal{B}$ and the same bounds c_1 and c_2 hold for f_k as for F, provided translation is taken into account. As in (5),

$$\sum_{x_i \in \mathcal{X}(\omega)} |f_k(x_i)|^2 < \frac{\|f_k\|_2^2}{k}.$$

Thus we have proved that, with probability 1, $\mathcal{X}$ fails to be a set of stable sampling. □

2.2. **Spatial Poisson processes.** Another scheme of choosing points randomly in $\mathbb{R}^d$ is the spatial Poisson process $\mathcal{X}$. This means that for some (intensity) function $\lambda : \mathbb{R}^d \to [0, \infty)$, for any Borel subset of $\mathbb{R}^d$, the number of points in $\mathcal{X} \cap A$ is a Poisson random variable with parameter $\int_A \lambda(x)\, dx$. If $A_1, \dots, A_n$ are disjoint sets, then the number of points in $\mathcal{X} \cap A_i$ are independent random variables.

The most natural case is where $\lambda(x)$ is a constant, $\lambda(x) = \rho$, say. Then the expected Beurling density of $\mathcal{X}$ is ρ. Again one might think that $\mathcal{X}$ is a set of stable sampling with high probability. However, as in Proposition 2.2, one cannot get the sampling inequality. In fact, a stronger result is true. One can choose points at a higher rate further from the origin and still have the sampling inequality failing.

Proposition 2.3. *Suppose $\mathcal{X}$ is a spatial Poisson process with $\lambda(x) = o(1 + \log^+(|x|))$. Then with probability one, the sampling inequality* (1) *fails for* every *subset $\mathcal{Y}$ of $\mathcal{X}$.*

Proof. Under the hypothesis on λ, the Beurling density of $\mathcal{X}$ may be infinite. In this case, $\mathcal{X}$ contains too many samples and the upper bound in the sampling inequality (1) will fail to hold. This problem could be fixed by extracting a subset

$\mathcal{Y}$ of $\mathcal{X}$ that satisfies the necessary conditions of Proposition 2.1. So one may still hope that a subsequence $\mathcal{Y}$ may yield a set of stable sampling.

However, we will show that with probability one, for each $k>0$ there exists a cube of side length k that contains no point of $\mathcal{X}$. Since the maximal hole of a set of stable sampling is bounded by Proposition 2.1(iii), with probability one, no subset of $\mathcal{X}$ can be a set of stable sampling.

The probability that a Poisson random variable with parameter λ is equal to zero is $e^{-\lambda}$. If X_i, $i=1,\dots,n$, are independent Poisson variables with parameters λ_i, resp., then

$$\begin{aligned}\mathbb{P}(\text{at least one} \quad X_i \quad \text{is zero}) &= 1-\mathbb{P}(X_1\neq 0,\dots,X_n\neq 0)\\ &= 1-\prod_{i=1}^n \mathbb{P}(X_i\neq 0) = 1-\prod_{i=1}^n(1-e^{-\lambda_i})\\ &= 1-\exp\Big(\sum_{i=1}^n \log(1-e^{-\lambda_i})\Big)\\ &\geq 1-\exp\Big(-\sum_{i=1}^n e^{-\lambda_i}\Big).\end{aligned}$$

Let $\epsilon>0$ be chosen later. Choose $m_0>k$ large so that $\lambda(x)\leq \epsilon\log(|x|)$ if $|x|>m_0$. For each $m\geq m_0$ we can find at least m disjoint cubes of side length k lying in $B(0,3mk)\setminus B(0,2mk)$; call them $C_{m1},\dots,C_{mm}$. The number of points in $\mathcal{X}$ lying in any one of the C_{mj} is a Poisson random variable with parameter less than $c_1\epsilon(\log m)k^d$. So by the above, the probability that at least one of the C_{mj} is empty is greater than

$$1-\exp\Big(-me^{-c_1\epsilon(\log m)k^d}\Big).$$

If we choose ϵ so that $c_1\epsilon k^d\leq 1/2$, then the above probability is greater than

$$1-\exp\Big(-m^{1/2}\Big),$$

which will be greater than $1/2$ if m is large enough.

Let D_m be the event that at least one of the cubes C_{mj}, $j=1,\dots,m$, is empty. For m large, $\mathbb{P}(D_m)\geq 1/2$, and the D_m are independent. So by the Borel-Cantelli lemma the event D_m happens for infinitely many m, with probability 1. In particular, there must be at least one cube of side length k with no points of $\mathcal{X}$ in it. □

On the other hand, the rate of growth $\log^+(|x|)$ is critical. If the intensity function λ grows faster than some multiple of $\log^+(|x|)$, then the random sequence $\mathcal{X}$ cannot have large holes.

Proposition 2.4. *Suppose $\mathcal{X}$ is a spatial Poisson process with intensity $\lambda(x)\geq c_0(1+\log^+(|x|))$ for all x. Fix $\alpha>0$. If $c_0\geq (d+1)/\alpha^d$, then with probability one, there exists $R>0$, such that every cube $\alpha k+[0,\alpha]^d$ for $\alpha|k|\geq R$ contains at least one point of $\mathcal{X}$.*

Proof. Let $\mathbb{S}_\alpha$ be the collection of all cubes of the form $\alpha k + [0,\alpha]^d, k \in \mathbb{Z}^d$. We will show that with probability one, all but finitely many cubes in $\mathbb{S}_\alpha$ contain at least one point of $\mathcal{X}$.

Let C_k be the event that the cube $A = \alpha k + [0,\alpha]^d$ contains no point of $\mathcal{X}$. If $\alpha|k| \geq N$, then $\lambda(x) \geq c_0 \log N$, and thus $\lambda(A) \geq c_0 \alpha^d \log N$. Thus for $\alpha|k| \geq N$, the probability that this cube is empty is

$$\mathbb{P}(C_k) = e^{-\lambda(A)} \leq e^{-c_0\alpha^d \log N} = N^{-c_0\alpha^d} .$$

If we choose $c_0\alpha^d \geq d+1$, then $\sum_{k\in\mathbb{Z}^d} \mathbb{P}(C_k) < \infty$. Then by the Borel-Cantelli lemma, the probability that infinitely many of the cubes are empty is 0. Therefore from some R on (depending on ω), all cubes in $\mathbb{S}_\alpha$ that are at least R from the origin are nonempty. □

3. A Positive Result: Relevant Sampling

The key to the arguments in Section 2 was that random sampling sets have either arbitrarily large holes or can be concentrated near the zeros of a band-limited function. In the former case we then constructed a class of functions whose main energy is concentrated on the "hole"; in the latter case we constructed a class of functions with prescribed zeros. These classes then violate the sampling inequality.

To obtain positive results we change the focus. Since for no reasonable random sampling set the norm equivalence (1) holds with positive probability for all band-limited functions, we will restrict the class of functions, for which we ask (1) to hold. The natural idea is to sample a given f in the region where a significant part of the energy is located. In other words, we sample in the region of relevant values.

This idea motivates the following definition. Let $C_R = [-R/2, R/2]^d$ be the cube of length R centered at the origin. Its volume is $\operatorname{vol} C_R = R^d$.

Definition 1. Fix a "large" number $R > 0$ and a "small" $\delta \in (0,1)$. Set

$$\widetilde{\mathcal{B}}(R,\delta) = \{f \in \mathcal{B} : \int_{C_R} |f(x)|^2\, dx \geq (1-\delta)\|f\|_2^2\} \tag{6}$$

and

$$\mathcal{B}(R,\delta) = \{f \in \mathcal{B} : \|f\|_2^2 = 1 \text{ and } \int_{C_R} |f(x)|^2\, dx \geq 1-\delta\} \tag{7}$$

Then $\mathcal{B}(R,\delta)$ is the subset of $\mathcal{B}$ consisting of those band-limited functions whose energy is largely concentrated on the cube C_R. Only a fraction δ of the total energy is outside this cube. It now makes sense to sample such f on the cube C_R and to expect that these samples are relevant and capture the main features of f.

Indeed, we will prove the following result.

Theorem 3.1. *Assume that $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are uniformly distributed over the cube $C_R = [-R/2, R/2]^d$ and $0 < \mu < 1-\delta$.*

Then there exist $A, B > 0$ such that the sampling inequality

$$\frac{r}{R^d}(1-\delta-\mu)\|f\|_2^2 \le \sum_{j=1}^{r} |f(x_j)|^2 \le \frac{r}{R^d}(1+\mu)\|f\|_2^2 \qquad \forall f \in \widetilde{\mathcal{B}}(R,\delta) \tag{8}$$

holds with probability at least

$$1 - 2Ae^{-B\frac{r}{R^d}\frac{\mu^2}{41+\mu}},$$

where B can be taken to be $B = \frac{1}{64\sqrt{2}}(e\log 2)^2 \approx 0.0392$. For large R and sufficiently large sampling size $r \ge C4^d R^{2d}/\mu^2$ with $C \approx 1.23 \cdot 10^4$, the constant A can be estimated explicitly by $A \approx \exp\left(\frac{16}{e}(4R)^d\right)$.

3.1. **Discussion and Open Problems.** 1. We emphasize that the exponential probability inequality holds uniformly for all $f \in \widetilde{\mathcal{B}}(R,\delta)$. By contrast, for fixed f such an inequality could be derived much more simply from standard limit theorems.

2. Theorem 3.1 is an asymptotic result. It is effective only for sufficiently large sampling sizes. In order to keep the constant A explicit as $A = \exp(\frac{16}{e}(4R)^d)$, we need $r \ge 2 \cdot 4^d R^{2d} \max\left(150.1/\mu, 56^2/\mu^2\right)$, see Section 4.4. In this case, we may ask how many samples are required to obtain (8) with "overwhelming probability". To achieve (8) with a probability exceeding $1-\epsilon$, we need $2Ae^{-B\frac{r}{R^d}\frac{\mu^2}{41+\mu}} < \epsilon$ or

$$r \ge \frac{R^d(41+\mu)}{B\mu^2}\Big(\log\frac{2}{\epsilon} + \frac{16}{e}(4R)^d\Big) \approx 1045\frac{R^d}{\mu^2}\Big(\log\frac{2}{\epsilon} + \frac{16}{e}(4R)^d\Big). \tag{9}$$

Since $\mathcal{B}(R,\delta)$ sits in a space of approximate dimension $D = R^d$, we need $\mathcal{O}(D^2)$ samples to recover every $f \in \mathcal{B}(R,\delta)$. In finite dimensional problems, for instance, when sampling trigonometric polynomials of fixed degree, one can often use random matrix techniques to show that the effective number of samples is in fact of the order $\mathcal{O}(D\log D)$ [13]. It is open whether this bound is achievable for band-limited functions in $\mathcal{B}(R,\delta)$.

3. The sampling inequality (8) states that every $f \in \mathcal{B}(R,\delta)$ is uniquely determined by a sufficient, but *finite* number of samples in $[-R/2, R/2]^d$. This may seem paradoxical at first glance, because the set of band-limited functions such that $f(x_j) = 0$ for $j = 1,\dots,r$, is an infinite-dimensional subspace of $\mathcal{B}$. However, as we assume that f is essentially supported on the cube $[-R/2, R/2]^d$, this means that f must take large values there. If $f(x_j) = 0$ for sufficiently many $x_j \in [-R/2, R/2]^d$, then f would oscillate and thus have a large derivative. But this would contradict the band-limitedness, which implies that the derivatives of f are bounded by π. While Theorem 3.1 is a probabilistic results, it seems possible to also prove a deterministic sampling inequality (8) for $\mathcal{B}(R,\delta)$, at least in dimension $d = 1$.

4. We emphasize that $\mathcal{B}(R,\delta)$ is not a subspace. This means that the frame algorithm (a linear reconstruction method) [12] cannot be used to recover f from its samples. Likewise, the projection-onto-convex-sets (POCS) method cannot be applied, because $\mathcal{B}(R,\delta)$ is not convex. Although (8) determines each $f \in \mathcal{B}(R,\delta)$

uniquely, currently we do not have an explicit reconstruction algorithm to recover f from its relevant samples.

4. Proof of Theorem 3.1

Theorem 3.1 will be a consequence of a large deviation inequality that holds uniformly over the whole class $\mathcal{B}(R,\delta)$ and will be proved in the following sections.

4.1. **Time-Frequency Limiting Operators.** Let P_R and Q be the projection operators defined by

$$P_R f = \chi_{C_R} f \qquad \text{and} \qquad Qf = \mathcal{F}^{-1}(\chi_{[-1/2,1/2]^d}\hat{f}) \,, \tag{10}$$

where $\mathcal{F}$ is the inverse Fourier transform. Then Q is the orthogonal projection from $L^2(\mathbb{R}^d)$ onto $\mathcal{B}$ and Q_R is the restriction of a function to the cube C_R. The composition

$$A_R = QP_RQ \tag{11}$$

is the operator of time and frequency limiting. This operator has been studied in detail by Landau, Slepian, Pollak [18, 19, 29–31] and many others. It encodes many deep properties of band-limited functions and their restrictions. In particular, A_R is a compact positive operator of trace class and a precisely known eigenvalue distribution.

We summarize the properties of the spectrum that will be needed in the sequel.

Let $A_R^{(1)}$ denote the operator of time-frequency limiting in dimension $d = 1$. Explicitly, $A_R^{(1)}$ is defined on $L^2(\mathbb{R})$ by the formula

$$(A_R^{(1)}\hat{f})(\xi) = \int_{-1/2}^{1/2} \frac{\sin \pi R(\xi-\eta)}{\pi(\xi-\eta)} \hat{f}(\eta)\, d\eta \qquad \text{for } |\xi| \leq 1/2\,.$$

We denote its eigenvalues by $\mu_k = \mu_k(R)$ in decreasing order and indicate the dependence on R. Then the first $[R]$ eigenvalues are approximately 1, followed by a "plunge region" of thickness $\mathcal{O}(\log R)$ after which the remaining eigenvalues are almost zero. Precisely, $\mu_{[R]+1}(R) \leq 1/2 \leq \mu_{[R]-1}(R)$ [16]. This behavior of the eigenvalues is usually formulated by saying that functions with spectrum $[-1/2, 1/2]$ and "essential" support on $[-R/2, R/2]$ form a finite-dimensional subspace of "approximate" dimension R. In particular, we may think of $\mathcal{B}(R,\delta)$ as a subset of a finite-dimensional space of dimension R.

The precise asymptotic behavior of the μ_k was obtained by Widom [33, Lemma 1-3]: he showed that for large k

$$\mu_k(R) \asymp 2\pi\Big(\frac{\pi R}{8}\Big)^{2k+1}\frac{1}{k!^2}\,, \tag{12}$$

where $a_k \asymp b_k$ means that $\lim_{k\to\infty} a_k/b_k = 1$. In particular, (12) implies the super-exponential decay

$$\mu_k(R) \leq C\exp\Big(-2k\log\big(\frac{2k}{\pi R}\big)\Big)\,. \tag{13}$$

We will use the following weaker exponential estimate.

Lemma 4.1. *[33] Given* $\alpha > 0$ *there exists a constant* $\kappa > 0$*, such that*

$$\mu_k(R) \le e^{-k/\kappa} \qquad for\, k \ge \frac{R}{1-\alpha}\,. \tag{14}$$

REMARK: This result is an asymptotic result for both $R \to \infty$ and $k \to \infty$. We emphasize that the constant κ depends only on α, but not on R. (Widom works with the operator $\int_{-1}^{1} \frac{\sin\gamma(\xi-\eta)}{\pi(\xi-\eta)} \hat{f}(\eta)\, d\eta$, so a simple dilation shows that we have to use $\gamma = \pi R/2$ to obtain $A_R^{(1)}$.)

Let $C(\epsilon)$ be the function counting the number of eigenvalues of $A_R^{(1)}$ exceeding ϵ, precisely

$$C(\epsilon) = \operatorname{card}\{\mu_k : \mu_k \ge \epsilon\}\,. \tag{15}$$

Then Lemma 4.1 implies that

$$C(\epsilon) \le \frac{R}{1-\alpha} + \kappa \log\frac{1}{\epsilon}\,. \tag{16}$$

A different estimate for the eigenvalue count was obtained by Landau and Widom [20]:

$$C(\epsilon) = R + \frac{2}{\pi}\log\frac{1-\epsilon}{\epsilon}\,\log R + o\Big(\log R\Big)\,. \tag{17}$$

However, this is an asymptotic result for $R \to \infty$, and its proof leaves open whether the term $o\big(\log R\big)$ can be chosen independent of ϵ. By contrast, the weaker estimate (16) works with a constant κ independent of R, at the price of the factor $(1-\alpha)^{-1}$. Since we need the eigenvalue behavior for *fixed* R, we use Widom's earlier result.

Next consider the time-frequency limiting operator A_R on $L^2(\mathbb{R}^d)$ Clearly A_R is the d-fold tensor product of $A_R^{(1)}$, $A_R = A_R^{(1)} \otimes \cdots \otimes A_R^{(1)}$. Consequently, λ is an eigenvalue of A_R, $\lambda \in \sigma(A_R)$, if and only if $\lambda = \prod_{j=1}^d \mu_{k_j}$, where $\mu_{k_j} \in \sigma(A_R^{(1)})$ is an eigenvalue of the one-dimensional operator $A_R^{(1)}$. Since $0 < \mu_k < 1$, we have $\prod_{j=1}^d \mu_{k_j} \ge \epsilon$ only when $\mu_{k_j} \ge \epsilon$ for $j = 1, \dots, d$. Consequently,

$$\{\lambda \in \sigma(A_R) : \lambda \ge \epsilon\} \subseteq \{\lambda = \prod_{j=1}^{d} \mu_{k_j} : \mu_{k_j} \in \sigma(A_R^{(1)}), \mu_k \ge \epsilon\}\,. \tag{18}$$

We arrange the eigenvalues of A_R by magnitude $1 > \lambda_1 \ge \lambda_2 \ge \lambda_3 \cdots \ge \lambda_n \ge \lambda_{n+1} \ge \cdots > 0$ and again let $C(\epsilon) = \max\{n : \lambda_n \ge \epsilon\}$ be the function counting the number of eigenvalues of A_R exceeding ϵ. We choose $\alpha = 1/2$ and combine (16) and (18); then the eigenvalue distribution for A_R in dimension d is

$$C(\epsilon) \le \Big(2R + \kappa \log\frac{1}{\epsilon}\Big)^d, \tag{19}$$

where κ is independent of R and ϵ. In the following, the decay constant κ is the only constant that is not explicit.

4.2. **Covering Number for $\mathcal{B}(R,\delta)$.** Recall that the covering numbers $N(\epsilon) = N(C,\epsilon)$ of a compact set C in a Banach space are defined to be the minimum number of balls of radius less than or equal to ϵ required to cover C. For the covering number of balls in Euclidean space we use a well-known estimate, see [7, 10].

Lemma 4.2. *Let $D(0,r) = \{x \in \mathbb{R}^d : \|x\|_2 \le r\}$ be the ball of radius r in $\mathbb{R}^d$. The covering number of $D(0,r)$ is given by*

$$N(\epsilon) = e^{d \log \frac{r}{\epsilon}} . \tag{20}$$

Let us note that the covering number of the shell $D(0,r) \setminus D(0,r(1-\delta))$ for some $\delta > 0$ is of the order $N(\epsilon) = e^{d\log\frac{r}{\epsilon}} - e^{d\log\frac{r(1-\delta)}{\epsilon}} = e^{d\log\frac{r}{\epsilon}}\big(1 - e^{d\log(1-\delta)}\big)$. The difference to the full ball is thus negligible for large dimensions.

Lemma 4.3. *(i) $\mathcal{B}(R,\delta)$ is a compact subset in $\mathcal{B}$.*

(ii) The covering number $N(\epsilon)$ of $\mathcal{B}(R,\delta)$ is bounded by

$$N(\epsilon) \le e^{2(2R+\kappa\log\frac{1}{\epsilon})^d \log(2/\epsilon)} . \tag{21}$$

Proof. (i) Clearly, $\mathcal{B}(R,\delta)$ is closed in $L^2(\mathbb{R}^d)$ and thus in $\mathcal{B}$ and consists of all $f \in \mathcal{B}(R,\delta)$ such that $\|f\|_2 = 1$ and $\langle A_R f, f\rangle \ge 1-\delta$. Therefore $\mathcal{B}(R,\delta)$ must be compact in $\mathcal{B}$.

(ii) Let φ_n be the eigenfunctions of A_R corresponding to the eigenvalues of λ_n. (These are tensor products of the standard prolate spheroidal functions.) Then $\{\varphi_n : n \in \mathbb{N}\}$ is an orthonormal basis for $\mathcal{B}$. If $f \in \mathcal{B}$, then $A_R f = \sum_{j=1}^\infty \lambda_n \langle f, \varphi_n\rangle \varphi_n$. Consequently $f \in \mathcal{B}(R,\delta)$ if and only if $\|f\|_2^2 = \sum_{n=1}^\infty |\langle f,\varphi_n\rangle|^2 = 1$ and $\langle A_R f, f\rangle = \sum_{n=1}^\infty \lambda_n |\langle f,\varphi_n\rangle|^2 \ge 1-\delta$. Define the subset of ℓ^2

$$S_\delta = \{\mathbf{c} \in \ell^2 : \|\mathbf{c}\|_2 = 1, \sum_{n=1}^\infty \lambda_n |c_n|^2 \ge 1-\delta\} .$$

Then $\mathcal{B}(R,\delta)$ and S_δ are isomorphic and their covering numbers are identical.

To cover S_δ with ϵ-balls, we split the constraint defining S_δ into two parts:

$$\sum_{n=1}^\infty \lambda_n |c_n|^2 = \sum_{\lambda_n \ge \epsilon/2} \lambda_n |c_n|^2 + \sum_{\lambda_n < \epsilon/2} \lambda_n |c_n|^2 = I_\epsilon + II_\epsilon \ge 1-\delta .$$

Since $II_\epsilon < \epsilon/2$, we must have $I_\epsilon \ge 1-\delta-\epsilon/2$, and it suffices to find the covering number for the set $S_\delta^I = \{\mathbf{c} : \|\mathbf{c}\|_2 \le 1, \sum_{\lambda_n \ge \epsilon/2} \lambda_n |c_n|^2 \ge 1-\delta-\epsilon/2\}$. Then S_δ^I is contained in the unit ball of a finite-dimensional Euclidean space $\mathbb{C}^N$. Its complex dimension is $N = C(\epsilon/2) = \max\{n \in \mathbb{N} : \lambda_n \ge \epsilon/2\}$ and its real dimension is $2C(\epsilon/2)$, where C is the counting function for the eigenvalues of A_R. By Lemma 4.2, S_δ^I can be covered by $\exp\Big(2C(\epsilon/2)\log(2/\epsilon)\Big)$ balls of radius $\epsilon/2$. Consequently, the infinite-dimensional set S_δ can be covered by $\exp\Big(2C(\epsilon/2)\log(2/\epsilon)\Big)$ ϵ-balls in ℓ^2 and we obtain the estimate

$$N(\epsilon) \le e^{2C(\epsilon/2)\log(2/\epsilon)} . \tag{22}$$

Using the eigenvalue distribution of Widom (Lemma 4.1) and (19) we obtain that

$$N(\epsilon) \le e^{2(2R+\kappa \log \frac{2}{\epsilon})^d \log \frac{2}{\epsilon}} .$$

□

REMARK: Estimate (22) also follows from general principles in approximation theory [24]. The estimate of the covering number by means of the eigenvalue distribution, equivalently between entropy numbers and approximation numbers, goes back to an inequality of Mityagin [22, Ch. 9].

A basic inequality for band-limited functions: *If $f \in \mathcal{B}$, then*

$$\|f\|_\infty \le \|f\|_2 \qquad \forall f \in \mathcal{B} . \tag{23}$$

This follows from the inversion formula $f(x) = \int_{[-1/2,1/2]^d} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi$ and Cauchy-Schwarz. As a consequence, every ϵ-net for $\mathcal{B}(R,\delta)$ in the L^2-norm is also an ϵ-net in the L^∞-norm.

For later use, we rewrite the covering number for $\epsilon = 2^{-\ell}$. Then we have

$$N(2^{-\ell}) \le \exp\Big(2\Big(2R + (\ell+1)\kappa \log 2\Big)^d (\ell+1)\log 2\Big) := \exp p(\ell) , \tag{24}$$

where $p(\ell) = 2\big(2R + (\ell+1)\kappa\log 2\big)^d (\ell+1)\log 2$ is a polynomial of degree $d+1$.

What is crucial in the above estimate, is that the exponent grows polynomially in ℓ, but not faster.

4.3. **Preparation for the proof of Theorem 3.1.** Assume that $\{x_j : j \in \mathbb{N}\}$ is an infinite sequence of i.i.d. random variables, each of which is uniformly distributed over the cube C_R.

For every $f \in \mathcal{B}$ we introduce the random variable

$$Y_j(f) = |f(x_j)|^2 - \frac{1}{R^d}\int_{C_R} |f(x)|^2\, dx = |f(x_j)|^2 - \mathbb{E}\left[\,|f(x_j)|^2\right] . \tag{25}$$

Then $Y_j(f)$ is a sequence of independent random variables with $\mathbb{E}\, Y_j(f) = 0$.

We first estimate the probability distribution of the random variable

$$\sup_{f\in\mathcal{B}(R,\delta)} \sum_{j=1}^r Y_j(f) .$$

For the repeated application of Bernstein's inequality for independent random variables we will need the following estimates for the $Y_j(f)$'s.

Lemma 4.4. *Let $f, g \in \mathcal{B}(R,\delta)$ and $j \in \mathbb{N}$. Then the following inequalities hold:*

$$\operatorname{Var} Y_j(f) \le \frac{1}{R^d} , \tag{26}$$

$$\operatorname{Var}\left(Y_j(f) - Y_j(g)\right) \le \frac{4}{R^d}\|f-g\|_\infty^2 , \tag{27}$$

$$\|Y_j(f)\|_\infty \le 1 , \tag{28}$$

$$\|Y_j(f) - Y_j(g)\|_\infty \le 2\|f-g\|_\infty . \tag{29}$$

Proof. We abbreviate the expected value of $|f(x_j)|^2$ by $\mu(f) = R^{-d}\int_{C_R}|f(x)|^2\,dx$. Using (23), we obtain

$$\begin{aligned}\operatorname{Var} Y_j(f) &= \mathbb{E}\,[Y_j(p)^2] = \mathbb{E}\,[\,|f(x_j)|^4] - \mu(f)^2 \\ &= \frac{1}{R^d}\int_{C_R}|f(x)|^4\,dx - \mu(f)^2 \\ &\le \frac{1}{R^d}\|f\|_\infty^2\,\|f\|_2^2 \le \frac{1}{R^d}\,.\end{aligned}$$

Similarly, we obtain

$$\|Y_j(p)\|_\infty = \sup_{\omega\in\Omega}\Big|\,|f(x_j(\omega))|^2 - \mu(f)\Big| \le \max\Big(\|f\|_\infty^2, \frac{1}{R^d}\int_{C_R}|f(x)|^2\,dx\Big) \le 1\,.$$

To prove (27), we write

$$\begin{aligned}\operatorname{Var}(Y_j(f) - Y_j(g)) &= \mathbb{E}\,(Y_j(f) - Y_j(g))^2 \\ &= \frac{1}{R^d}\int_{C_R}\big(|f(x)|^2 - |g(x)|^2\big)^2\,dx - (\mu(f) - \mu(g))^2 \\ &\le \frac{1}{R^d}\int_{C_R}|f(x) - g(x)|^2\,\big(|f(x)| + |g(x)|\big)^2\,dx \\ &\le \frac{2}{R^d}\|f - g\|_\infty^2\int_{\mathbb{R}^d}(|f(x)|^2 + |g(x)|^2)\,dx \le \frac{4}{R^d}\|f - g\|_\infty^2\end{aligned}$$

The last estimate follows similarly from

$$\begin{aligned}\|Y_j(f) - Y_j(g)\|_\infty &\le \sup_{\omega\in\Omega}\Big(\Big|\,|f(x_j(\omega))|^2 - |g(x_j(\omega))|^2\Big| - \frac{1}{R^d}\int_{C_R}(|f(x)|^2 - |g(x)|^2)\,dx\Big) \\ &\le \||f|^2 - |g|^2\|_\infty \\ &\le \|f - g\|_\infty\,\|\,|f| + |g|\|_\infty \\ &= 2\|f - g\|_\infty\,.\end{aligned}$$

□

4.4. **Proof of the sampling inequality.** The sampling inequality follows from a uniform large deviation inequality for the sampling of band-limited functions.

Theorem 4.5. *Let $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are uniformly distributed over $C_R = [-R/2, R/2]^d$. Then there exist constants $A, B > 0$ depending on d and R, such that*

$$\mathbb{P}\Big(\sup_{f\in\mathcal{B}(R,\delta)}\Big|\sum_{j=1}^r Y_j(f)\Big| \ge \lambda\Big) \le 2A\exp\Big(-B\frac{\lambda^2}{41rR^{-d} + \lambda}\Big) \tag{30}$$

for $r \in \mathbb{N}$ and $\lambda \ge 0$.

Here $B = \frac{1}{64\sqrt{2}}(e\log 2)^2 \approx 0.0392$. If R is sufficiently large, for instance $R \ge 2\kappa(d+1)/e$ where κ is Widom's constant, and if $\lambda \ge 150.1(4R)^d + 78.5\cdot 2^d\sqrt{r}$, then (30) holds with $A = \exp\Big(\frac{16}{e}(4R)^d\Big)$.

Before we prove the large deviation inequality, we show how the main theorem follows from Theorem 4.5.

Proof of Theorem 3.1. Choose $\lambda = \frac{r\mu}{R^d}$ and recall that $Y_j(f) = |f(x_j)|^2 - R^{-d}\int_{C_R} |f(x)|^2\,dx$. Thus the event $\mathcal{E} = \{\sup_{f\in\mathcal{B}(R,\delta)} |\sum_{j=1}^s Y_j(f)| \geq r\mu R^{-d}\}$ coincides with the event

(31)
$$\frac{r}{R^d}\int_{C_r} |f(x)|^2\,dx - \frac{r\mu}{R^d} \leq \sum_{j=1}^r |f(x_j)|^2 \leq \frac{r}{R^d}\int_{C_r} |f(x)|^2\,dx + \frac{r\mu}{R^d} \qquad \text{for } all\ f \in \mathcal{B}(R,\delta)\,.$$

Since by definition $1-\delta \leq \int_{C_R} |f(x)|^2\,dx \leq 1$, we find that the event of the uniform sampling inequality

$$\frac{r(1-\mu-\delta)}{R^d} \leq \sum_{j=1}^r |f(x_j)|^2 \leq \frac{r(1+\mu)}{R^d} \qquad \text{for all } f \in \mathcal{B}(R,\delta) \tag{32}$$

is contained in $\mathcal{E}$. As a consequence of Theorem 4.5 the sampling inequality (32) holds uniformly for all $f \in \mathcal{B}(R,\delta)$ with probability at least

$$1 - 2A\exp(-BrR^{-d}\mu^2/(41+\mu)).$$

Finally, we need $\lambda = r\mu/R^d \geq 150.1(4R)^d + 78.5 \cdot 2^d\sqrt{r}$. This requires about $r \geq 2\cdot 4^d R^{2d} \max\big(150.1/\mu, 78.5^2/\mu^2\big) \approx 1.23\cdot 10^4 4^d R^{2d}/\mu^2$ samples.

This proves Theorem 3.1. □

We are left to prove the probability estimate of Theorem 4.5. To estimate the probability of the deviation of a sum of random variables from its average we use Bernstein's inequality for independent random variables [2]: *Let $Y_j, j = 1,\dots,r$, be a sequence of bounded, independent random variables with* $\mathbb{E}\,Y_j = 0$, $\mathrm{Var} Y_j \leq \sigma^2$, *and $\|Y_j\|_\infty \leq M$ for $j = 1,\dots,r$. Then*

$$\mathbb{P}\Big(\Big|\sum_{j=1}^r Y_j\Big| \geq \lambda\Big) \leq 2\exp\Big(-\frac{\lambda^2}{2r\sigma^2 + \frac{2}{3}M\lambda}\Big)\,. \tag{33}$$

Proof of Theorem 4.5. Step 1: *A metric entropy argument.* For a given $\ell \in \mathbb{N}$, we construct an $2^{-\ell}$-covering for $\mathcal{B}(R,\delta)$ with respect to the L^2-norm. Let $\mathcal{A}_\ell$ be the corresponding $2^{-\ell}$-net. Then $\bigcup_{h\in\mathcal{A}_\ell} B(h,2^{-\ell}) \supseteq \mathcal{B}(R,\delta)$ and $\mathcal{A}_\ell$ can be chosen to have cardinality at most $N(2^{-\ell})$ (the covering number of $\mathcal{B}(R,\delta)$). Note that by (23) this covering is also a $2^{-\ell}$-covering for the supremum norm.

Given $f \in \mathcal{B}(R,\delta)$, let f_ℓ be the function in $\mathcal{A}(2^{-\ell})$ that is closest to f in L^∞-norm, with some convention for breaking ties. Since $\|f - f_\ell\|_2 \to 0$ and f_0 can be chosen to be $f_0 = 0$, we can write

$$Y_j(f) = (Y_j(f_1) - Y(f_0)) + (Y_j(f_2) - Y_j(f_1)) + (Y_j(f_3) - Y_j(f_2)) + \cdots\,.$$

If $\sup_{f\in\mathcal{B}(R,\delta)} |\sum_{j=1}^r Y_j(f)| \geq \lambda$, then

$$\text{there exist } \ell \geq 1, f \in \mathcal{A}(2^{-\ell}),\ \text{ and } g \in \mathcal{A}(2^{-\ell+1}) \ \text{ with } \|f-g\|_\infty \leq 3\cdot 2^{-\ell},$$

$$\text{such that } \Big|\sum_{j=1}^r \big(Y_j(f) - Y_j(g)\big)\Big| \geq \lambda/2\ell^2\,. \tag{34}$$

If this were not the case, then

$$\Big|\sum_{j=1}^r Y_j(f)\Big| \le \sum_{\ell=1}^\infty \Big|\sum_{j=1}^r (Y_j(f_\ell)-Y_j(f_{\ell-1}))\Big| \le \sum_{\ell=1}^\infty \frac{\lambda}{2\ell^2} = \frac{\pi^2}{12}\lambda < \lambda.$$

Let $\mathcal{E}_\ell$ be the event defined in (34). Next we estimate the probability of $\mathcal{E}_\ell$ for each $\ell \ge 1$.

Step 2. We estimate the term $\ell = 1$ separately, since $\mathcal{A}(0) = \{0\}$ and so $g = 0$. For fixed $f \in \mathcal{A}(1/2)$, the probability of the event in (a) is bounded, using Bernstein's inequality (33) and Lemma 4.4, by

$$2\exp\Big(-\frac{\lambda^2/4}{2r\mathrm{Var}\, Y_j(f) + \frac{2}{3}(\lambda/2)\|Y_j(f)\|_\infty}\Big) \le 2\exp\Big(-\frac{\lambda^2}{2rR^{-d}+\lambda/3}\Big).$$

There are at most $N(1/2) = \exp\big(2(2R+2\kappa\log 2)^d 2\log 2\big)$ functions in $\mathcal{A}(1/2)$, so the probability of $\mathcal{E}_1$ is bounded by

$$2\exp\big(2(2R+2\kappa\log 2)^d 2\log 2\big)\,\exp\Big(-\frac{\lambda^2}{2rR^{-d}+\lambda/3}\Big). \tag{35}$$

Step 3. For $\ell \ge 2$, we estimate the probability of $\mathcal{E}_\ell$ in a similar fashion by using Lemma 4.4, (27), and (29). If $f \in \mathcal{A}(2^{-\ell})$ and $g \in \mathcal{A}(2^{-\ell+1})$ with $\|f-g\|_\infty \le 3\cdot 2^{-\ell}$, we have

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{j=1}^r (Y_j(f) \;&-\; Y_j(g))\Big| > \frac{\lambda}{2\ell^2}\Big)\\ &\le\; 2\exp\Big(-\frac{\lambda^2/4\ell^4}{2r\cdot 4\cdot R^{-d}(3\cdot 2^{-\ell})^2 + \frac{2}{3}2\cdot 3\cdot 2^{-\ell-1}\lambda/\ell^2}\Big)\\ &=\; 2\exp\Big(-\frac{2^\ell}{8\ell^2}\,\frac{\lambda^2}{36rR^{-d}\ell^2 2^{-\ell}+\lambda}\Big).\end{aligned}$$

Note $36\ell^2/2^\ell < 41$. There are at most $N(2^{-\ell})$ functions in $\mathcal{A}(2^{-\ell})$ and $N(2^{-\ell+1})$ functions in $\mathcal{A}(2^{-\ell+1})$. Finally, this can happen for any ℓ. So the probability of $\bigcup_{\ell=2}^\infty \mathcal{E}_\ell$ is bounded by

$$\begin{aligned}\sum_{\ell=2}^\infty \quad & N(2^{-\ell})N(2^{-\ell+1})2\exp\Big(-\frac{2^\ell}{8\ell^2}\,\frac{\lambda^2}{41rR^{-d}+\lambda}\Big)\\ \le\; & \sum_{\ell=2}^\infty 2\exp\Big(p(\ell)+p(\ell-1)-\frac{2^\ell}{8\ell^2}\,\frac{\lambda^2}{41rR^{-d}+\lambda}\Big),\end{aligned} \tag{36}$$

where we use (24) for the covering number.

Step 4. We will need the following inequality:

If $p, a > 0$, *then*

$$\sum_{\ell=2}^\infty e^{-a^\ell p} \le \frac{1}{pa\log a} e^{-ap}. \tag{37}$$

This inequality follows from the integral test and the substitution $a^x = u$:

$$\begin{aligned}\sum_{\ell=2}^{\infty} e^{-a^\ell p} &\le \int_1^\infty e^{-a^x p} dx \\ &= \frac{1}{\log a}\int_a^\infty e^{-pu}\,\frac{du}{u} \\ &\le \frac{1}{a\log a}\int_a^\infty e^{-pu}\,du \\ &= \frac{1}{pa\log a} e^{-ap}.\end{aligned}$$

Step 5. To estimate the sum (36), we rewrite and simplify each term. Set

$$\psi = \frac{\lambda^2}{41rR^{-d}+\lambda} \tag{38}$$

$$c_1 = \min_{\ell\ge 2} \frac{2^{\ell/2}}{8\ell^2} \tag{39}$$

$$c_2 = \max_{\ell\ge 2} \frac{2p(\ell)}{2^{\ell/2}}. \tag{40}$$

and $\sup_{\ell\ge 2} \ell^2/2^\ell = 9/8$. Then the ℓ-th term in (36) is majorized by

$$\exp\Big(-2^{\ell/2}(c_1\psi - c_2)\Big).$$

If $\psi > 0$ is large enough so that $p := c_1\psi - c_2 > 0$, then (37) implies that

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{\ell=2}^{\infty}\mathcal{E}_\ell\Big) &\le 2\frac{1}{(c_1\psi-c_2)\sqrt{2}\log\sqrt{2}}\, e^{-\sqrt{2}(c_1\psi-c_2)} \\ &= \frac{2\sqrt{2}}{\log 2}\,\frac{e^{\sqrt{2}c_2}}{c_1\psi - c_2}\,\exp\Big(-\frac{\sqrt{2}c_1\lambda^2}{41rR^{-d}+\lambda}\Big). \qquad (41)\end{aligned}$$

Since the term for $\ell = 1$ has the same form, we have proved that

$$\mathbb{P}\Bigg(\sup_{f\in\mathcal{B}(R,\delta)}\Big|\sum_{j=1}^{r} Y_j(f)\Big| \ge \lambda\Bigg) \le 2A\exp\Bigg(-\frac{\sqrt{2}c_1\lambda^2}{41rR^{-d}+\lambda}\Bigg),$$

whenever $\psi > c_2/c_1$.

For the exponent B we may take the smaller of the exponents in (35) and (41), i.e., $B = \min(3, \sqrt{2}c_1)$. If we choose λ large enough, so that $c_1\psi - c_2 \ge \frac{2\sqrt{2}}{\log 2}$, then we may take $A = \max(\exp\big(2(2R + 2\kappa\log 2)^d 2\log 2\big), e^{\sqrt{2}c_2})$.

Step 6. Finally we estimate the constants c_1 and c_2, A and B. Let $f_a(x) = x^a/2^{x/2}$ for $a > 0, x > 0$. An elementary calculation shows that $\max_{x\ge 0} f_a(x) = \Big(\frac{2a}{e\log 2}\Big)^a$. Consequently

$$c_1 = \frac{1}{8}\min_{\ell\ge 2}\frac{2^{\ell/2}}{\ell^2} \ge \frac{1}{8}\Big(\frac{e\log 2}{4}\Big)^2,$$

so the exponent B in (30) is $\sqrt{2}c_1 = \frac{1}{64\sqrt{2}}(e\log 2)^2 \approx 0.0392$.

As for c_2, recall that $p(\ell) = 2\big(2R + (\ell+1)\kappa\log 2\big)^d(\ell+1)\log 2$. Either we have $(\ell+1)\kappa\log 2 \le 2R$, in which case

$$c_2 \le 4(4R)^d \max_{\ell\ge 2} \frac{\ell+1}{2^{\ell/2}} \log 2 \le 2^{5/2}(4R)^d \frac{2}{e\log 2} \log 2 = \frac{2^{7/2}}{e}(4R)^d .$$

Else $(\ell+1)\kappa\log 2 > 2R$, so that

$$\begin{aligned} c_2 &\le 4\big(2(\ell+1)\kappa\log 2\big)^d(\ell+1)\log 2/2^{\ell/2} \\ &\le 2(2\log 2)^{d+1}\sqrt{2}\,\kappa^d \max_{\ell\ge 2} \frac{(\ell+1)^{d+1}}{2^{(\ell+1)/2}} \\ &\le 2^{3/2}(2\log 2)^{d+1}\,\kappa^d\Big(\frac{2(d+1)}{e\log 2}\Big)^{d+1} \\ &= 2^{3/2}\,\kappa^d 4^{d+1}\Big(\frac{d+1}{e}\Big)^{d+1} . \end{aligned}$$

Thus we find that

$$c_2 \le \max\Big(\frac{2^{7/2}}{e}(4R)^d, 2^{3/2}\,\kappa^d 4^{d+1}\Big(\frac{d+1}{e}\Big)^{d+1}\Big) . \tag{42}$$

So for R large enough, roughly $R \ge \kappa(d+1)/e$, we may choose A to be $A = e^{\sqrt{2}c_2} = \exp\Big(\frac{16}{e}(4R)^d\Big)$.

Finally consider the condition $c_1\psi - c_2 \ge \frac{2^{3/2}}{\log 2}$ or equivalently, $\frac{\lambda^2}{41rR^{-d}+\lambda} \ge \frac{c_2+2^{3/2}/\log 2}{c_1} := C \approx 150.1\,(4R)^d$. Since $x \ge B + \sqrt{D}$ implies $x^2 \ge Bx + D$, we find that

$$\lambda \ge C + (41CrR^{-d})^{1/2} \approx 150.1(4R)^d + 78.5\cdot 2^d\sqrt{r} .$$

□

References

[1] R. F. Bass and K. Gröchenig. Random sampling of multivariate trigonometric polynomials. *SIAM J. Math. Anal.*, 36(3):773–795 (electronic), 2004/05.

[2] G. Bennett. Probability inequalities for the sum of independent random variables. *Amer. Stat. Ass. J.*, 57:33–45, 1962.

[3] A. Beurling. Local harmonic analysis with some applications to differential operators. In *Some Recent Advances in the Basic Sciences, Vol. 1 (Proc. Annual Sci. Conf., Belfer Grad. School Sci., Yeshiva Univ., New York, 1962–1964)*, pages 109–125. Belfer Graduate School of Science, Yeshiva Univ., New York, 1966.

[4] R. P. Boas, Jr. *Entire functions*. Academic Press, New York, 1954.

[5] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[6] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[7] B. Carl and I. Stephani. *Entropy, compactness and the approximation of operators*, volume 98 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1990.

[8] G. Chistyakov and Y. Lyubarskii. Random perturbations of exponential Riesz bases in $L^2(-\pi,\pi)$. *Ann. Inst. Fourier (Grenoble)*, 47(1):201–255, 1997.
[9] G. Chistyakov, Y. Lyubarskii, and L. Pastur. On completeness of random exponentials in the Bargmann-Fock space. *J. Math. Phys.*, 42(8):3754–3768, 2001.
[10] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39(1):1–49 (electronic), 2002.
[11] R. M. Dudley. Sample functions of the Gaussian process. *Ann. Probability*, 1(1):66–103, 1973.
[12] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72:341–366, 1952.
[13] K. Gröchenig, B. Pötscher, and H. Rauhut. Learning trigonometric polynomials from random samples and exponential inequalities for eigenvalues of random matrices. *Preprint*, 2007.
[14] K. Gröchenig and H. Razafinjatovo. On Landau's necessary density conditions for sampling and interpolation of band-limited functions. *J. London Math. Soc. (2)*, 54(3):557–565, 1996.
[15] A. Iosevich and M. N. Kolountzakis. A Weyl type formula for Fourier spectra and frames. *Proc. Amer. Math. Soc.*, 134(11):3267–3274 (electronic), 2006.
[16] H. Landau. On the density of phase space expansions. *IEEE Trans. Inform. Theory*, 39:1152–1156, 1993.
[17] H. J. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.*, 117:37–52, 1967.
[18] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. II. *Bell System Tech. J.*, 40:65–84, 1961.
[19] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. III. The dimension of the space of essentially time- and band-limited signals. *Bell System Tech. J.*, 41:1295–1336, 1962.
[20] H. J. Landau and H. Widom. Eigenvalue distribution of time and frequency limiting. *J. Math. Anal. Appl.*, 77(2):469–481, 1980.
[21] B. Y. Levin. *Lectures on entire functions.* American Mathematical Society, Providence, RI, 1996. In collaboration with and with a preface by Yu. Lyubarskii, M. Sodin and V. Tkachenko, Translated from the Russian manuscript by Tkachenko.
[22] G. G. Lorentz. *Approximation of functions.* Holt, Rinehart and Winston, New York, 1966.
[23] S. Mendelson and A. Pajor. On singular values of matrices with independent rows. *Bernoulli*, 12(5):761–773, 2006.
[24] A. Pinkus. *n-widths in approximation theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin, 1985.
[25] T. Poggio and S. Smale. The mathematics of learning: dealing with data. *Notices Amer. Math. Soc.*, 50(5):537–544, 2003.
[26] D. Potts, G. Steidl, and M. Tasche. Fast Fourier transforms for nonequispaced data: a tutorial. In *Modern sampling theory*, Appl. Numer. Harmon. Anal., pages 247–270. Birkhäuser Boston, Boston, MA, 2001.
[27] K. Seip. *Interpolation and sampling in spaces of analytic functions*, volume 33 of *University Lecture Series.* American Mathematical Society, Providence, RI, 2004.
[28] K. Seip and A. M. Ulanovskii. Random exponential frames. *J. London Math. Soc. (2)*, 53(3):560–568, 1996.
[29] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainity. IV. Extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Tech. J.*, 43:3009–3057, 1964.
[30] D. Slepian. On bandwidth. *Proc. IEEE*, 64(3):292–300, 1976.
[31] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. I. *Bell System Tech. J.*, 40:43–63, 1961.
[32] S. Smale and D.-X. Zhou. Shannon sampling and function reconstruction from point values. *Bull. Amer. Math. Soc. (N.S.)*, 41(3):279–305 (electronic), 2004.

[33] H. Widom. Asymptotic behavior of the eigenvalues of certain integral equations. II. *Arch. Rational Mech. Anal.*, 17:215–229, 1964.

Department of Mathematics, The University of Connecticut, Storrs, CT 06269-3009, USA

Faculty of Mathematics, University of Vienna, Nordbergstrasse 15, A-1090 Vienna, Austria

E-mail address: bass@math.uconn.edu,karlheinz.groechenig@univie.ac.at