RELEVANT SAMPLING OF BAND-LIMITED FUNCTIONS

RICHARD F. BASS AND KARLHEINZ GRÖCHENIG

Abstract. We study the random sampling of band-limited functions of several variables. If a band-limited function with bandwidth has its essential support on a cube of volume $R^d$, then $O(R^d \log R^d)$ random samples suffice to approximate the function up to a given error with high probability.

1. Introduction

The nonuniform sampling of band-limited functions of several variables remains a challenging problem. Whereas in dimension 1 the density of a set essentially characterizes sets of stable sampling [14], in higher dimensions the density is no longer a decisive property of sets of stable sampling. Only a few strong and explicit sufficient conditions are known, e.g., [3, 10, 12].

This difficulty is one of the reasons for taking a probabilistic approach to the sampling problem [2,20]. At first glance, one would guess that every reasonably homogeneous set of points in $\mathbb{R}^d$ satisfying Landau’s necessary density condition will generate a set of stable sampling. This intuition is far from true. To the best of our knowledge, every construction in the literature of sets of random points in $\mathbb{R}^d$ contains either arbitrarily large holes with positive probability or concentrates near the zero manifold of a band-limited function. Both properties are incompatible with a sampling inequality. See [2] for a detailed discussion.

The difficulties with the probabilistic approach lie in the unboundedness of the configuration space $\mathbb{R}^d$ and the infinite dimensionality of the space of band-limited functions. To resolve this issue, we argued in [2] that usually one observes only finitely many samples of a band-limited function and that these observations are drawn from a bounded subset of $\mathbb{R}^d$. Moreover, since it does not make sense to sample a given function $f$ in a region where $f$ is small, we proposed to sample $f$ only on its essential support. Since $f$ is sampled only in the relevant region, this method might be called the “relevant sampling of band-limited functions.” In this paper we continue our investigation of the random sampling of band-limited functions and settle a question that was left open in [2], namely how many random samples are required to approximate a band-limited function locally to within a given accuracy?

2000 Mathematics Subject Classification. 94A20, 42C15, 60E15, 62M30.

K. G. was supported in part by the project P22746-N13 of the Austrian Science Foundation (FWF).

R. B. was supported in part by NSF grant DMS-0901505.
To fix terms, recall that 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 \( \hat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-2\pi i x \cdot \xi} \, dx \). A set \( \{ x_j : j \in J \} \subseteq \mathbb{R}^d \) is called a set of stable sampling or simply a set of sampling [7], if there exist constants \( A, B > 0 \), such that a sampling inequality holds:
\[
(1) \quad A\|f\|_2^2 \leq \sum_{j} |f(x_j)|^2 \leq B\|f\|_2^2, \quad \forall f \in \mathcal{B}.
\]

Next, we sample only on the essential support of \( f \). Therefore we let \( C_R = [-R/2, R/2]^d \) and define the subset
\[ \mathcal{B}(R, \delta) = \left\{ f \in \mathcal{B} : \int_{C_R} |f(x)|^2 \, dx \geq (1 - \delta)\|f\|_2^2 \right\}. \]

As a continuation of [2], we will prove the following sampling theorem.

**Theorem 1.** Let \( \{ x_j : j \in \mathbb{N} \} \) be a sequence of independent and identically distributed random variables that are uniformly distributed in \( C_R \). Suppose that \( R \geq 2 \), that \( \delta \in (0, 1) \) and \( \nu \in (0, 1/2) \) are small enough, and that \( 0 < \epsilon < 1 \). There exists a constant \( \kappa \) so that if the number of samples \( r \) satisfies
\[
(2) \quad r \geq R^d \frac{1 + \nu/3}{\nu^2} \log \frac{2R^d}{\epsilon},
\]
then the sampling inequality
\[
(3) \quad \frac{r}{R^d} \left( \frac{1}{2} - \delta - \nu - 12\delta\kappa \right)\|f\|_2^2 \leq \sum_{j=1}^{r} |f(x_j)|^2 \leq r\|f\|_2^2 \quad \text{for all } f \in \mathcal{B}(R, \delta)
\]
holds with probability at least \( 1 - \epsilon \). The constant \( \kappa \) can be taken to be \( \kappa = e^{d\pi} \).

The formulation of Theorem 1 is similar to [2, Thm. 3.1]. The main point is that only \( O(R^d \log R^d) \) samples are required for a sampling inequality to hold with high probability. In [2] we used a metric entropy argument to show that \( O(R^{2d}) \) samples suffice. We expect that the order \( O(R^d \log R^d) \) is optimal. We point out that in addition all constants are now explicit.

Our idea is to replace the sampling of band-limited function in \( \mathcal{B}(R, \delta) \) by a finite-dimensional problem, namely the sampling of the corresponding span of prolate spheroidal functions on the cube \( [-R/2, R/2]^d \) and then use error estimates. For the probability estimates we use a new tool, namely the powerful matrix Bernstein inequality of Ahlswede and Winter [1] in the optimized version of Tropp [22].

The remainder of the paper contains the analysis of a related finite-dimensional problem for prolate spheroidal functions in Section 2 and transition to the infinite-dimensional problem in \( \mathcal{B}(R, \delta) \) with the necessary error estimates in Section 3. The appendix contains an elementary estimate for the constant \( \kappa \).
2. Finite-Dimensional Subspaces of $\mathcal{B}$

We first study a sampling problem in a finite-dimensional subspace related to the set $\mathcal{B}(R, \delta)$.

**Prolate Spheroidal Functions.** Let $P_R$ and $Q$ be the projection operators defined by

\begin{equation}
P_R f = \chi_{C_R} f \quad \text{and} \quad Q f = \mathcal{F}^{-1}(\chi_{[-1/2,1/2]^d} \hat{f}),
\end{equation}

where $\mathcal{F}^{-1}$ is the inverse Fourier transform. The composition of these orthogonal projections

\begin{equation}
A_R = Q P_R Q
\end{equation}

is the operator of time and frequency limiting. This operator arises frequently in the context of band-limited functions and uncertainty principles. The localization operator $A_R$ is a compact positive operator of trace class, and by results of Landau, Slepian, Pollak, and Widom \[8,9,17,19,23\] the eigenvalue distribution spectrum is precisely known. We summarize the properties of the spectrum that we will need.

Let $A^{(1)}_R$ denote the operator of time-frequency limiting in dimension $d = 1$. This operator can be defined explicitly on $L^2(\mathbb{R})$ by the formula

\begin{equation}
(A^{(1)}_R f) \hat{\gamma}(\xi) = \int_{-1/2}^{1/2} \sin \pi R (\xi - \eta) \hat{f} (\eta) d\eta \quad \text{for } |\xi| \leq 1/2.
\end{equation}

The eigenfunctions of $A^{(1)}_R$ are the prolate spheroidal functions and let the corresponding eigenvalues $\mu_k = \mu_k(R)$ be arranged in decreasing order. According to \[6\] they satisfy

\begin{align*}
0 < \mu_k(R) < 1 & \quad \forall k \in \mathbb{N}, \\
\mu_{[R]+1}(R) \leq 1/2 & \leq \mu_{[R]-1}(R); \\
\mu_{[R]+1}(R) \leq 1/2 & \leq \mu_{[R]-1}(R);
\end{align*}

As a consequence any function with spectrum $[-1/2,1/2]$ and “essential” support on $[-R/2,R/2]$ is close to the span of the first $R$ prolate spheroidal functions. In particular, we may think of $\mathcal{B}(R, \delta)$ as, roughly, almost a subset of a finite-dimensional space of dimension $R$.

The time-frequency limiting operator $A_R$ on $L^2(\mathbb{R}^d)$ is the $d$-fold tensor product of $A^{(1)}_R$, $A_R = A^{(1)}_R \otimes \cdots \otimes A^{(1)}_R$. Consequently, $\sigma(A_R)$, the spectrum of $A_R$, is

$$
\sigma(A_R) = \{ \lambda \in (0, 1) : \lambda = \prod_{j=1}^d \mu_{k_j}, \mu_{k_j} \in \sigma(A^{(1)}_R) \}.
$$

Since $0 < \mu_k < 1$, $A_R$ possesses at most $R^d$ eigenvalues greater than or equal to 1/2. Again we arrange the eigenvalues of $A_R$ by magnitude $1 > \lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_n \geq \lambda_{n+1} \geq \cdots > 0$. Let $\phi_j$ be the eigenfunction corresponding to $\lambda_j$.

We fix $R$ “large” and $\delta \in (0, 1)$. Let

$$
\mathcal{P}_N = \text{span} \{ \phi_j : j = 1, \ldots, N \}
$$

be a finite-dimensional subspace of $\mathcal{B}(R, \delta)$.
be the span of the first \( N \) eigenfunctions of the time-frequency limiting operator \( A_R \) (one might call them “multivariate prolate polynomials”). For properly chosen \( N \), \( P_N \) consists of functions in \( \mathcal{B}(R, \delta) \). See Lemma 5.

By Plancherel’s theorem,
\[
\langle Qf, g \rangle = \langle \hat{f}, \chi_{[-1/2,1/2]} \hat{g} \rangle = \langle f, Qg \rangle.
\]
Then for \( f \in \mathcal{B} \) we have \( Qf = f \), and so
\[
\langle A_R f, f \rangle = \langle P_R Qf, Qf \rangle = \langle P_R f, f \rangle = \int_{C_R} |f(x)|^2 dx.
\]

We first study random sampling in the finite-dimensional space \( P_N \). In the following \( \|f\|_{2,R} \) denotes the normalized \( L^2 \)-norm of \( f \) restricted to the cube \( C_R = [-R/2, R/2]^d \):
\[
\|f\|_{2,R}^2 = \int_{C_R} |f(x)|^2 dx.
\]

**Proposition 2.** Let \( \{x_j : j \in \mathbb{N}\} \) be a sequence of independent and identically distributed random variables that are uniformly distributed in \( [-R/2, R/2]^d \). Then
\[
\mathbb{P} \left( \inf_{f \in P_N, \|f\|_{2,R}=1} \sum_{j=1}^r \left( |f(x_j)|^2 - \frac{1}{R^d} \|f\|_{2,R}^2 \right) \leq - \frac{\nu}{R^d} \right) \leq N \exp \left( - \frac{\nu^2 r}{R^d (1 + \nu/3)} \right)
\]
for \( r \in \mathbb{N} \) and \( \nu \geq 0 \).

**Proof.** We prove the proposition in several steps. First, since \( P_N \) is finite-dimensional, the sampling inequality for \( P_N \) amounts to a statement about the spectrum of an underlying (random) matrix.

Let \( f = \langle c, \phi \rangle = \sum_{k=1}^N c_k \phi_k \in P_N \), so that \( |f(x_j)|^2 = \sum_{k,l=1}^N c_k \bar{c}_l \phi_k(x_j) \phi_l(x_j) \). Now define the \( N \times N \) matrix \( T_j \) of rank one by letting the \( (k, l) \) entry be
\[
(T_j)_{kl} = \phi_k(x_j) \phi_l(x_j).
\]
Then \( |f(x_j)|^2 = \langle c, T_j c \rangle \). Since each random variable \( x_j \) is uniformly distributed over \( C_R \) and \( \phi_k \) is the \( k \)-th eigenfunction of the localization operator \( A_R \), using (6) the expectation of the \( kl \)-th entry is
\[
E \left( (T_j)_{kl} \right) = \frac{1}{R^d} \int_{C_R} \phi_k(x) \phi_l(x) dx = \frac{1}{R^d} \langle A_R \phi_k, \phi_l \rangle = \frac{1}{R^d} \delta_{kl} \phi_k(x_j) \phi_l(x_j),
\]
where \( \delta_{kl} \) is Kronecker’s delta. Consequently the expectation of \( T_j \) is the diagonal matrix
\[
E(T_j) = \frac{1}{R^d} \text{diag} (\lambda_k) =: \frac{1}{R^d} \Delta.
\]
We may now rewrite the expression in (7) as
\[
\inf_{f \in \mathcal{P}, \|f\|_2 = 1} \frac{1}{r} \sum_{j=1}^{r} \left( |f(x_j)|^2 - \frac{1}{R^d} \|f\|_{2,R}^2 \right)
\]
\[
= \inf_{\|c\|_2 = 1} \frac{1}{r} \sum_{j=1}^{r} \left( \langle c, T_j c \rangle - \langle c, \mathbb{E}(T_j)c \rangle \right)
\]
\[
= \lambda_{\min} \left( \frac{1}{r} \sum_{j=1}^{N} (T_j - \mathbb{E}(T_j)) \right)
\]
(11)

where we use \(\lambda_{\min}(U)\) for the smallest eigenvalue of a self-adjoint matrix \(U\).

Consequently, we have to estimate a probability for the matrix norm of a sum of random matrices. We do this using a matrix Bernstein inequality due to Tropp [22]. Let \(\lambda_{\max}(A)\) be the largest singular value of a matrix \(A\) so that \(\|A\| = \lambda_{\max}(A^*A)^{1/2}\) is the operator norm (with respect to the \(\ell^2\)-norm).

**Theorem 3.** (Tropp) Let \(X_j\) be a sequence of independent, random self-adjoint \(N \times N\)-matrices. Suppose that
\[
\mathbb{E}X_j = 0 \quad \text{and} \quad \|X_j\| \leq B \quad \text{a.s.}
\]
and let
\[
\sigma^2 = \left\| \sum_{j=1}^{r} \mathbb{E}(X_j^2) \right\|.
\]
Then for all \(t \geq 0\),
\[
\mathbb{P} \left( \lambda_{\max} \left( \sum_{j=1}^{r} X_j \right) \geq t \right) \leq N \exp \left( - \frac{t^2/2}{\sigma^2 + Bt/3} \right).
\]

To apply the matrix Bernstein inequality, we set \(X_j = T_j - \mathbb{E}(T_j)\). We need to calculate \(\|X_j\|\) and \(\|\sum_j \mathbb{E}(X_j^2)\|\). Clearly \(\mathbb{E}(X_j) = 0\).

**Lemma 4.** Under the conditions stated above we have
\[
\|X_j\| \leq 1,
\]
\[
\mathbb{E}(X_j^2) \leq R^{-d} \Delta,
\]
and
\[
\sigma^2 = \left\| \sum_{j=1}^{r} \mathbb{E}(X_j^2) \right\| \leq \frac{r}{R^d}.
\]

**Proof.** (i) To estimate the matrix norm of \(X_j\), recall that
\[
\|X_j\| = \sup_{\|f\|_2 = 1} \left| |f(x_j)|^2 - R^{-d} \|f\|_{2,R}^2 \right| \leq \|f\|_\infty - R^{-d} \|f\|_{2,R}^2 \leq \|f\|_2 = 1.
\]

Hence we obtain
\[
\|X_j\| = \sup_{\|f\|_2 = 1} \left| |f(x_j)|^2 - R^{-d} \|f\|_{2,R}^2 \right| \leq \|f\|_\infty - R^{-d} \|f\|_{2,R}^2 \leq \|f\|_2 = 1.
\]
(ii) Next we calculate the matrix $\mathbb{E}(X_j^2)$:

$$
\mathbb{E}(X_j^2) = \mathbb{E}(T_j^2) - R^{-d}\mathbb{E}(T_j\Delta) + R^{-2d}\frac{\Delta^2}{d}
$$

Furthermore, the square of the rank one matrix $T_j$ is the (rank one) matrix

$$(T_j^2)_{km} = \sum_{l=1}^{N} \langle \phi_l(x_j), T_j \rangle \langle \phi_l(x_j), \phi_m(x_j) \rangle
$$

Writing $m(x) = \sum_{l=1}^{N} |\phi_l(x)|^2$, we obtain

$$
T_j^2 = m(x)T_j
$$

Let $s$ be the function whose Fourier transform is given by $\hat{s} = \chi_{[-1/2,1/2]^d}$ and let $T_x f(t) = f(t - x)$ be the translation operator. Then it is well known that $T_x s$ is the reproducing kernel for $\mathcal{B}$, that is,

$$
f(x) = \langle f, T_x s \rangle.
$$

To see this, by Plancherel’s theorem and the inversion formula for the Fourier transform, if $f \in \mathcal{B}$,

$$
\langle f, T_x s \rangle = \langle \hat{f}, e^{-2\pi i x \cdot \xi} \hat{s} \rangle = \int_{[-1/2,1/2]^d} e^{2\pi i x \cdot \xi} \hat{f}(\xi) \, d\xi = \int e^{2\pi i x \cdot \xi} \hat{f}(\xi) \, d\xi = f(x).
$$

Since the eigenfunctions $\phi_l$ form an orthonormal basis for $\mathcal{B}$, the factor $m(x_j)$ in (14) is majorized by

$$
m(x_j) = \sum_{l=1}^{N} |\phi_l(x_j)|^2 \leq \sum_{l=1}^{N} |\langle \phi_l, T_x s \rangle|^2 = \sum_{l=1}^{\infty} |\langle \phi_l, T_x s \rangle|^2 = \|T_x s\|_2^2 = 1.
$$

Since $T_j^2 \leq T_j$ and the expectation preserves the cone of positive (semi)definite matrices (see, e.g. [22]), we have $\mathbb{E}(T_j^2) \leq \mathbb{E}(T_j) = R^{-d}\Delta$, and

$$
\mathbb{E}(X_j^2) = \mathbb{E}(T_j^2) - R^{-2d}\frac{\Delta^2}{d} \leq R^{-d}\Delta.
$$

(iii) Now the variance of the sum of positive (semi)definite random matrices is majorized by

$$
\sigma^2 = \left\| \sum_{j=1}^{r} \mathbb{E}(X_j^2) \right\| \leq \left\| \sum_{j=1}^{r} \mathbb{E}(T_j) \right\| = \frac{r}{R^d} \|\Delta\| \leq \frac{r}{R^d}.
$$
End of the proof of Proposition 2. Now we have all information to finish the proof of Proposition 2. Since $\lambda_{\min}(T) = -\lambda_{\max}(-T)$, we substitute these estimates into the matrix Bernstein inequality with $t = r\nu/R^d$, and obtain that

$$
\mathbb{E}\left(\lambda_{\min}\left(\sum_{j=1}^{r} (T_j - \mathbb{E}(T_j))\right) \leq -r\nu/R^d\right) \leq N \exp\left(-\frac{r^2 \nu^2 R^{-2d}}{rR^{-d} + r\nu R^{-d}/3}\right).
$$

Combined with (11), the proposition is proved. \qed

Random matrix theory offers several methods to obtain probability estimates for the spectrum of random matrices. In [2] we used the entropy method. We also mention the influential work of Rudelson [15] and the recent papers [11,16] on random matrices with independent columns. The matrix Bernstein inequality offers a new approach and makes the probabilistic part of the argument almost painless. The matrix Bernstein inequality was first derived in [1] and improved in several subsequent papers, in particular in [13]. The version with the best constants is due to Tropp [22]. Matrix Bernstein inequalities also simplify many probabilistic arguments in compressed sensing; see the forthcoming book [4].

3. From Sampling of Prolate Spheroidal Functions to Relevant Sampling of Bandlimited Functions

Let $\alpha$ be the value of the $N$-th eigenvalue of $A_R$, that is, $\alpha = \lambda_N$, let $E = E_N$ be the orthogonal projections from $B$ onto $P_N$, and let $F = F_N = I - E_N$. Intuitively, since $f \in \mathcal{B}(R, \delta)$ is essentially supported on the cube $C_R$, it should be close to the span of the largest eigenfunctions of $A_R$ and thus $Ff$ should be small. The following lemma gives a precise estimate. Compare also with the proof of [9, Thm. 3].

Lemma 5. If $f \in \mathcal{B}(R, \delta)$, then

$$
\|Ef\|_2^2 \geq \left(1 - \frac{\delta}{1 - \alpha}\right) \|f\|_2^2,
$$

$$
\|Ef\|_{2,R}^2 \geq \alpha \left(1 - \frac{\delta}{1 - \alpha}\right) \|f\|_2^2,
$$

$$
\|Ff\|_2^2 \leq \frac{\delta}{1 - \alpha} \|f\|_2^2.
$$

Proof. Expand $f \in \mathcal{B}$ with respect to the prolate spheroidal functions as $f = \sum_{j=1}^{\infty} c_j \phi_j$. Without loss of generality, we may assume that $\|f\|_2 = \|c\|_2 = 1$. Since $f \in \mathcal{B}(R, \delta)$, we have that

$$
1 - \delta \leq \|f\|_{2,R}^2 = \int_{C_R} |f(t)|^2 dt = \langle A_R f, f \rangle = \sum_{j=1}^{\infty} |c_j|^2 \lambda_j.
$$

Set

$$
A = \|Ef\|_2^2 = \sum_{j=1}^{N} |c_j|^2
$$

and

$$
\lambda_{\min}(T) = -\lambda_{\max}(-T),
$$

we have

$$
\lambda_{\min}(\sum_{j=1}^{r} (T_j - \mathbb{E}(T_j))) \leq -r\nu/R^d \leq N \exp\left(-\frac{r^2 \nu^2 R^{-2d}}{rR^{-d} + r\nu R^{-d}/3}\right).
$$

Combined with (11), the proposition is proved. \qed
and $B = \sum_{j > N} |c_j|^2 = 1 - A = \|F f\|_2^2$. Since $\lambda_j \leq \lambda_N = \alpha$ for $j > N$, we estimate $A = \|E f\|_2^2$ as follows:

$$A = \sum_{j=1}^N |c_j|^2 \geq \sum_{j=1}^N \lambda_j |c_j|^2 \geq 1 - \delta - \lambda_N \sum_{j=N+1}^\infty |c_j|^2 = 1 - \delta - \alpha (1 - A).$$

The inequality $A \geq 1 - \delta - \alpha (1 - A)$ implies that $\|E f\|_2^2 = A \geq 1 - \frac{\delta}{1 - \alpha}$ and using the orthogonal decomposition $f = Ef + Ff$,

$$B = \|F f\|_2^2 \leq \frac{\delta}{1 - \alpha}.$$

Finally, $\|E f\|_{2,R}^2 = \sum_{j=1}^N \lambda_j |c_j|^2 \geq \alpha A \geq \alpha (1 - \frac{\delta}{1 - \alpha})$, as claimed.

**REMARK** (due to J.-L. Romero): As mentioned in [2], if $f \in \mathcal{B}(R, \delta)$ and $f(x_j) = 0$ for sufficiently many samples $x_j \in C_R$, then $f \equiv 0$. However, $f$ cannot be completely determined by samples in $C_R$ alone. This is a consequence of the fact that $\mathcal{B}(R, \delta)$ is not a linear space. Given a finite subset $S \subseteq C_R$, consider the finite-dimensional subspace $\mathcal{H}_0$ of $\mathcal{B}$ spanned by the reproducing kernels $T_{x}s, x \in S$. If $\phi \in \mathcal{H}_0^\perp$, then $\phi(x) = \langle \phi, T_{x}s \rangle = 0$ for $x \in S$. Thus by adding a function in $\mathcal{H}_0^\perp$ of sufficiently small norm to $f \in \mathcal{B}(R, \delta)$, one obtains a different function with the same samples. More precisely, let $f \in \mathcal{B}(R, \delta)$ with $\|f\|_2 = 1$ and $\int_{C_R} |f(x)|^2 dx = \gamma > 1 - \delta$ and $\phi \in \mathcal{H}_0^\perp$ with $\|\phi\|_2 = 1$. Then $f(x) + \epsilon \phi(x) = f(x)$ for $x \in S$ and $f + \epsilon \phi \in \mathcal{B}(R, \delta)$ for sufficiently small $\epsilon > 0$.

Despite this non-uniqueness, one can approximate $f$ from the samples up to an accuracy $\delta$, as is shown by the next lemma.

We will require a standard estimate for sampled 2-norms, a so-called Plancherel-Polya-Nikolskij inequality [21]. Assume that $\mathcal{X} = \{x_j\} \subseteq \mathbb{R}^d$ is relatively separated, i.e., the “covering index”

$$\max_{k \in \mathbb{Z}^d} \# \mathcal{X} \cap (k + [-1/2, 1/2]^d) =: N_0 < \infty$$

is finite. Then there exists a constant $\kappa > 0$, such that

$$\sum_{j=1}^\infty |f(x_j)|^2 \leq \kappa N_0 \|f\|_2^2 \quad \text{for all } f \in \mathcal{B}.$$

The constant $\kappa$ can be chosen as $\kappa = e^{\bar{d} \pi}$. Since the standard proof in [21] uses a maximal inequality with an non-explicit constant, we will give a simple argument using Taylor series in the appendix.
Lemma 6. Let \( \{ x_j : j = 1, \ldots, r \} \) be a finite subset of \( C_R \) with covering index \( N_0 \). Then the solution to the least square problem

\[
(p_{opt} = \arg\min_{p \in P_N} \left\{ \sum_{j=1}^{r} |f(x_j) - p(x_j)|^2 \right\})
\]

satisfies the error estimate

\[
\sum_{j=1}^{r} |f(x_j) - p_{opt}(x_j)|^2 \leq N_0 \kappa \frac{\delta}{1-\alpha} \|f\|_2^2
\]

for all \( f \in \mathcal{B}(R, \delta) \).

Proof. We combine Lemma 5 with (15).

Next we compare sampling inequalities for the space of prolate polynomials \( P_N \) to sampling inequalities for functions in \( \mathcal{B}(R, \delta) \).

Lemma 7. Let \( \{ x_j : j = 1, \ldots, r \} \) be a finite subset of \( C_R \) with covering index \( N_0 \). If the inequality

\[
(18) \quad \frac{1}{r} \sum_{j=1}^{r} \left( |p(x_j)|^2 - R^{-d} \|p\|_{2,R}^2 \right) \geq - \frac{\nu}{R^d} \|p\|_2^2
\]

holds for all \( p \in P_N \), then the inequality

\[
(19) \quad \sum_{j=1}^{r} |f(x_j)|^2 \geq A \|f\|_2^2
\]

holds for all \( f \in \mathcal{B}(R, \delta) \) with a constant

\[
A = \frac{r}{R^d} \left( \alpha - \frac{\alpha \delta}{1-\alpha} - \nu \right) - 2 \kappa N_0 \frac{\delta}{1-\alpha}
\]

REMARK: For \( A \) to be positive we need

\[
r \geq R^d \frac{2 \kappa N_0 \frac{\delta}{1-\alpha}}{\alpha - \frac{\alpha \delta}{1-\alpha} - \nu}.
\]
Proof. Using the triangle inequality and the orthogonal decomposition $f = Ef + Ff$, we estimate

$$
\left( \sum_{j=1}^{r} |f(x_j)|^2 \right)^{1/2} \geq \left( \sum_{j=1}^{r} |Ef(x_j)|^2 \right)^{1/2} - \left( \sum_{j=1}^{r} |Ff(x_j)|^2 \right)^{1/2}.
$$

Taking squares and using (15) on $Ef$ and $Ff$ in the cross product term, we continue as

$$
\sum_{j=1}^{r} |f(x_j)|^2 \geq \sum_{j=1}^{r} |Ef(x_j)|^2 - 2 \left( \sum_{j=1}^{r} |Ef(x_j)|^2 \right)^{1/2} \left( \sum_{j=1}^{r} |Ff(x_j)|^2 \right)^{1/2}
$$

$$
+ \sum_{j=1}^{r} |Ff(x_j)|^2
$$

$$
\geq \sum_{j=1}^{r} |Ef(x_j)|^2 - 2\kappa N_0 \|Ef\|_2 \|Ff\|_2
$$

$$
\geq \sum_{j=1}^{r} |Ef(x_j)|^2 - 2\kappa N_0 \frac{\delta}{1 - \alpha} \|f\|_2^2,
$$

since by Lemma 5 $\|Ff\|_2^2 \leq \frac{\delta}{1 - \alpha} \|f\|_2^2$ and $\|Ef\|_2 \leq \|f\|_2$. Now we make use of hypothesis (18) and Lemma 5 and obtain

$$
\sum_{j=1}^{r} |f(x_j)|^2 \geq \sum_{j=1}^{r} |Ef(x_j)|^2 - 2\kappa N_0 \frac{\delta}{1 - \alpha} \|f\|_2^2
$$

$$
\geq \frac{r}{R^d} \|Ef\|_{2,R}^2 - \frac{\nu r}{R^d} \|Ef\|_2^2 - 2\kappa N_0 \frac{\delta}{1 - \alpha} \|f\|_2^2
$$

$$
\geq \frac{\alpha r}{R^d} \left( 1 - \frac{\delta}{1 - \alpha} \right) \|f\|_2^2 - \frac{\nu r}{R^d} \|f\|_2^2 - 2\kappa N_0 \frac{\delta}{1 - \alpha} \|f\|_2^2.
$$

So we may choose $A$ to be

$$
A = \frac{r}{R^d} \left( \alpha - \frac{\alpha \delta}{1 - \alpha} - \nu \right) - 2\kappa N_0 \frac{\delta}{1 - \alpha}.
$$

The final ingredient we need is a deviation inequality for the covering index $N_0 = \max_{k \in \mathbb{Z}^d} \{ x_j \} \cap (k + [-1/2, 1/2]^d)$.

Lemma 8. Suppose $R \geq 2$ and $\{ x_j : j = 1, \ldots, r \}$ are independent and identically distributed random variables that are uniformly distributed over $C_R$. Let $a > R^{-d}$. Then

$$
\mathbb{P}(N_0 > ar) \leq (R + 2)^d \exp \left( -r \left( a \log(aR^d) - (a - R^{-d}) \right) \right).
$$
Proof. Let $D_k = k + [-1/2, 1/2]^d$ for $k \in \mathbb{Z}^d$. Note that we need at most $(R + 2)^d$ of the $D_k$’s to cover $C_R$. If $N_0 > ar$, then for at least one $k$, $D_k$ must contain at least $ar$ of the $x_j$’s. Therefore

\[ P(N_0 > ar) \leq (R + 2)^d \max_{k \in \mathbb{Z}^d} P(\#\{x_j\} \cap D_k > ar). \]

Fix $k \in \mathbb{Z}^d$. For any $b > 0$, by Chebyshev’s inequality

\[ P(\#\{x_j\} \cap D_k > ar) = P\left(\sum_{j=1}^{r} \chi_{D_k}(x_j) > ar\right) = P\left(\exp\left(b \sum_{j=1}^{r} \chi_{D_k}(x_j)\right) > e^{bar}\right) \leq e^{-bar} \exp\left(b \sum_{j=1}^{r} \chi_{D_k}(x_j)\right). \]

Since the $x_j$ are uniformly distributed over $C_R$, then $\chi_{D_k}(x_j)$ is equal to 1 with probability at most $R^{-d}$ and otherwise equals zero. Therefore, using the independence,

\[ P(\#\{x_j\} \cap D_k > ar) \leq e^{-bar} \prod_{j=1}^{r} E e^{b\chi_{D_k}(x_j)} \leq e^{-bar}((1 - R^{-d}) + e^{-d} R^{-d})^r = e^{-bar}((1 + (e^{-d} - 1) R^{-d})^r \leq e^{-bar}(\exp((e^{-d} - 1) R^{-d}))^r. \]

With the optimal choice $b = \log(aR^d)$ the last term is then

\[ \exp\left(-r (a \log(aR^d) - (a - R^{-d}))\right). \]

Substituting this in (20) proves the lemma.

By combining the finite-dimensional result of Proposition 2 with the estimates of Lemmas 7 and 8 and the appropriate choice of the free parameters, we obtain the following theorem.

**Theorem 9.** Let $\{x_j : j \in \mathbb{N}\}$ be a sequence of independent and identically distributed random variables that are uniformly distributed in $C_R$. Suppose $R \geq 2$,

\[ \delta < \frac{1}{2(1 + 12\kappa)} \]

and

\[ \nu < \frac{1}{2} - \delta(1 + 12\kappa). \]

Let

\[ A = \frac{r}{R^d} \left(\frac{1}{2} - \delta - \nu - 12\delta\kappa\right). \]

Then the sampling inequality

\[ A \|f\|_2^2 \leq \sum_{j=1}^{r} |f(x_j)|^2 \leq r \|f\|_2^2 \quad \text{for all } f \in B(R, \delta) \]
holds with probability at least

\[ 1 - R^d \exp \left( -\frac{\nu^2 r}{R^d (1 + \nu/3)} \right) - (R + 2)^d \exp \left( -\frac{r}{R^d} (3 \log 3 - 2) \right). \]

**Proof.** Since \( |f(x)| \leq \|f\|_2 \) for \( f \in B \), the right hand inequality in (22) is immediate. We take \( \alpha = 1/2 \) and \( N = R^d \) in Proposition II and \( a = 3R^{-d} \) in Lemma 8. Let

\[ V_1 = \left\{ \inf_{f \in P_N \|f\|_2 = 1} \frac{1}{r} \sum_{j=1}^r (|f(x_j)|^2 - \frac{1}{R^d} \|f\|_{2,R}^2) \leq -\frac{\nu}{R^d} \right\} \]

and let \( V_2 = \{ N_0 > ar \} \).

By Proposition II and Lemma 8, the probability of \((V_1 \cup V_2)^c\) is bounded below by (23). By Lemma 7,

\[ \frac{1}{r} \sum_{j=1}^r |f(x_j)|^2 \geq A \|f\|^2 \]

for all \( f \in B(R, \delta) \) on the set \((V_1 \cup V_2)^c\). With \( \alpha = 1/2 \) and \( N_0 = 3R^{-d} \) the lower bound \( A \) of Lemma 7 simplifies to \( A = \frac{r}{R^d} \left( \frac{1}{2} - \delta - \nu - 12\delta \kappa \right) \). Our assumptions on \( \delta \) and \( \nu \) guarantee that \( A > 0 \).

The formulation of Theorem I now follows. With \( N = R^d \) and \( 0 < \nu < 1/2 - \delta < 1/2 \), if \( \epsilon > 0 \) is given and

\[ r \geq \max \left( R^d \frac{1 + \nu/3}{\nu^2} \log \frac{2R^d}{\epsilon}, \frac{R^d}{3 \log 3 - 2} \log \frac{2(R + 2)^d}{\epsilon} \right) = R^d \frac{1 + \nu/3}{\nu^2} \log \frac{2R^d}{\epsilon}, \]

then the probability in (23) will be larger than \( 1 - \epsilon \).

**Remark:** Observe that the parameters \( \delta \) and \( R \) are not independent. As mentioned in [2, p. 14], for \( B(R, \delta) \) to be non-empty, we need \( \delta \geq 2\pi \sqrt{2Re^{-\pi R}} \) (up to terms of higher order). Thus for small \( \delta \) as in Theorem 9 we need to choose \( R \) of order \( R \approx c \log(d/\delta) \).

**Appendix A. The Plancherel-Polya Inequality**

We finish by showing that the constant \( \kappa \) in the Plancherel-Polya inequality (15) can be chosen explicitly to be \( \kappa = e^{d\pi} \). The argument is simple and well-known, see, for example, [5].

**Lemma 10.** Let \( \{x_j : j \in \mathbb{N}\} \) be a set in \( \mathbb{R}^d \) with covering index \( N_0 \). Then

\[ \sum_{j=1}^\infty |f(x_j)|^2 \leq N_0 e^{d\pi} \|f\|^2. \]
Proof. Let $k \in \mathbb{Z}^d$ and $x_j \in k + [-1/2, 1/2] =: D_k$. Then $\|x_j - k\|_\infty \leq 1/2$. Consider the Taylor expansion of $f(x_j)$ at $k$ (with the usual multi-index notation):

$$|f(x_j)| = \left| \sum_{\alpha \geq 0} \frac{D^\alpha f(k)}{\alpha!} (x_j - k)^\alpha \right| \leq \sum_{\alpha \geq 0} \left| \frac{D^\alpha f(k)}{\alpha!} \right| \left( \frac{1}{2} \right)^{\alpha |\alpha|}.$$

We now let $\theta \in (0, 1)$ and apply Cauchy-Schwarz:

$$|f(x_j)|^2 \leq \sum_{\alpha \geq 0} \frac{1}{\alpha!} \left( \frac{1}{2} \right)^{2\theta |\alpha|} \sum_{\alpha \geq 0} \left| \frac{D^\alpha f(k)}{\alpha!} \right|^2 \left( \frac{1}{2} \right)^{2(1-\theta) |\alpha|} = e^{d/4^\theta} \sum_{\alpha \geq 0} \left| \frac{D^\alpha f(k)}{\alpha!} \right|^2 \left( \frac{1}{2} \right)^{2(1-\theta) |\alpha|}.$$

If $f \in \mathcal{B}$, then by Shannon’s sampling theorem (or because the reproducing kernels $T_k$, $k \in \mathbb{Z}^d$, form an orthonormal basis of $\mathcal{B}$) we have

$$\sum_{k \in \mathbb{Z}^d} |f(k)|^2 = \|f\|^2_2 \quad \forall f \in \mathcal{B}.$$

To estimate the partial derivatives we use Bernstein’s inequality $\|D^\alpha f\|_2 \leq \pi^{|\alpha|} \|f\|_2$.

We first assume that $N_0 = 1$, i.e., each cube $D_k$ contains at most one of the $x_j$’s. Then we obtain, after interchanging the order of summation

$$\sum_{j \in \mathbb{N}} |f(x_j)|^2 \leq e^{d/4^\theta} \sum_{\alpha \geq 0} \sum_{k \in \mathbb{Z}^d} \left| \frac{D^\alpha f(k)}{\alpha!} \right|^2 \left( \frac{1}{2} \right)^{2(1-\theta) |\alpha|}$$

$$= e^{d/4^\theta} \sum_{\alpha \geq 0} \left( \frac{1}{2} \right)^{2(1-\theta) |\alpha|} \|D^\alpha f\|_2^2 \left/ \alpha! \right.$$ 

$$\leq e^{d/4^\theta} \sum_{\alpha \geq 0} \left( \frac{1}{2} \right)^{2(1-\theta) |\alpha|} \pi^{|\alpha|} \|f\|^2_2 = e^{d/4^\theta} e^{d\pi^2/4^{1-\theta}} \|f\|^2_2$$

(26)

The choice $4^\theta = 2/\pi$ yields the constant $\kappa = e^{d/4^\theta} e^{d\pi^2/4^{1-\theta}} = e^{d\pi}$. For arbitrary $N_0$ we obtain

$$\sum_{j \in \mathbb{N}} |f(x_j)|^2 = \sum_{k \in \mathbb{Z}^d} \sum_{\{j : x_j \in D_k\}} |f(x_j)|^2 \leq N_0 e^{d\pi} \|f\|^2_2,$$

as claimed.

Possibly the Plancherel-Polya inequality could be improved to a local estimate of the form $\sum_{x_j \in C_R} |f(x_j)|^2 \leq \tilde{\kappa} N_0 \|f\|^2_{2,R}$, but we did not pursue this question.

**References**

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, 2002.

[2] R.F. Bass and K. Gröchenig. Random sampling of band-limited functions. *Israel J. Math.*, 177:1–28, 2010.
[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] S. Foucart and H. Rauhut. A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2012, to appear.

[5] K. Gröchenig. Reconstruction algorithms in irregular sampling. Math. Comp., 59(199):181–194, 1992.

[6] H. Landau. On the density of phase space expansions. IEEE Trans. Inform. Theory, 39:1152–1156, 1993.

[7] H. J. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. Acta Math., 117:37–52, 1967.

[8] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. II. Bell System Tech. J., 40:65–84, 1961.

[9] 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.

[10] B. Matei and Y. Meyer. Simple quasicrystals are sets of stable sampling. Complex Var. Elliptic Equ., 55(8-10):947–964, 2010.

[11] S. Mendelson and A. Pajor. On singular values of matrices with independent rows. Bernoulli, 12(5):761–773, 2006.

[12] A. Olevskii and A. Ulanovskii. Universal sampling and interpolation of band-limited signals. Geom. Funct. Anal., 18(3):1029–1052, 2008.

[13] R. I. Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. Electron. Commun. Probab., 15:203–212, 2010.

[14] J. Ortega-Cerdà and K. Seip. Fourier frames. Ann. of Math. (2), 155(3):789–806, 2002.

[15] M. Rudelson. Random vectors in the isotropic position. J. Funct. Anal., 164(1):60–72, 1999.

[16] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. Comm. Pure Appl. Math., 62(12):1707–1739, 2009.

[17] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty. IV. Extensions to many dimensions; generalized prolate spheroidal functions. Bell System Tech. J., 43:3009–3057, 1964.

[18] D. Slepian. On bandwidth. Proc. IEEE, 64(3):292–300, 1976.

[19] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. I. Bell System Tech. J., 40:43–63, 1961.

[20] 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.

[21] H. Triebel. Theory of function spaces. Birkhäuser Verlag, Basel, 1983.

[22] J. Tropp. User-friendly tail bounds for sums of random matrices. Found. Comput. Math., 2011. doi:10.1007/s10208-011-9099-z.

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

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

E-mail address: r.bass@uconn.edu

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

E-mail address: karlheinz.groechenig@univie.ac.at