RANDOM SAMPLING OF MULTIVARIATE TRIGONOMETRIC POLYNOMIALS

RICHARD F. BASS AND KARLHEINZ GRÖCHENIG

ABSTRACT. We investigate when a trigonometric polynomial p of degree M in d variables is uniquely determined by its sampled values $p(x_j)$ on a random set of points x_j in the unit cube (the "sampling problem for trigonometric polynomials") and estimate the probability distribution of the condition number for the associated Vandermonde-type and Toeplitz-like matrices. The results provide a solid theoretical foundation for some efficient numerical algorithms that are already in use.

1. INTRODUCTION

The reconstruction, interpolation or approximation of a function (signal, image) from a given data set is a central task in many problems of data processing. The mathematical problem is to find a function $f(x)$ in a suitable function space V that interpolates or approximates the given data $y_j = f(x_j)$. The set $\mathcal{X} = \{x_j : j = 1, \dots, r\} \subseteq \mathbb{R}^d$ is the sampling set, and the function space V comes from the mathematical modeling of signals or images (e.g., band-limitedness, smoothness). The numerical and theoretical analysis of the sampling problem depends, of course, heavily on the signal model V.

In this paper we focus almost exclusively on multivariate trigonometric polynomials as our model. While this is by no means the only possible model, it is convenient, interesting, and occurs in many applications where standard uniform sampling is not possible. Specifically, the model of trigonometric polynomials has been used in cardiology (one-dimensional) [37], geophysics (2-dimensional) [29], image processing (2-dimensional) [35], as a non-uniform discrete Fourier transform (1- and 2-dimensional) [8, 13, 14, 28, 33] and in computer tomography (2 and 3-dimensional) [3, 27, 32]. Furthermore the space of trigonometric polynomials of fixed degree is the appropriate finite-dimensional model for the approximation of band-limited functions from a finite number of samples [19, 20].

Clearly, the sampling operator $f \to \{f(x_j) : j = 1, \dots, r\}$ is linear, and, for a finite-dimensional model space, it can therefore be described by a matrix. For the model of trigonometric polynomials of fixed degree, this matrix possesses an

1991 *Mathematics Subject Classification.* 94A12,94A20,15A12,15A52,42B99,42A15,42A61, 60G50,60G99.

Key words and phrases. Sampling, band-limited functions, multivariate trigonometric polynomials, random sampling, block Toeplitz matrix, Vandermonde matrix, condition number, metric entropy.

The first author was partially supported by NSF grant DMS-0244737.

additional structure, namely it is either a rectangular Vandermonde-like matrix or a square Toeplitz-like matrix. This structure is the basis for efficient and fast numerical algorithms. For dimension $d = 1$ we refer to [8, 15, 16, 30, 36], for higher dimensions to [27,29,32,35]. These algorithms are fast, stable, and robust, but only in dimension $d = 1$ do the numerical algorithms possess a solid theoretical basis (invertibility, estimates of condition numbers and rates of convergence for iterative algorithms).

In higher dimensions, there is only numerical evidence that the existing algorithms work; except for some isolated results [18, 22] there has been no theoretical justification for the success of these numerical methods. The main reason for this disparity lies in the nature of zero sets of trigonometric polynomials in one and higher dimensions. In dimension $d = 1$ the zero set of a trigonometric polynomial is finite by the fundamental theorem of algebra, whereas the zero set of a trigonometric polynomial in several variables is an algebraic variety. This difference makes it almost impossible to determine effectively whether the reconstruction problem $\{f(x_j)\} \to f$ is solvable for a fixed multi-dimensional sampling set $\mathcal{X} \subseteq \mathbb{R}^d$. It seems even more difficult to estimate the condition numbers of the associated matrices. On the other hand, numerical experiments and successful applications make it plausible that for generic sampling sets $\mathcal{X} \subseteq \mathbb{R}^d$ the sampling problem is solvable and well-conditioned.

Our goal is to achieve some understanding for the success of existing numerical methods and to provide more insight into the theoretical issues. To do this we adopt a probabilistic point of view: Instead of seeking analytic statements for a fixed sampling set, we consider the collection of all sampling sets of size r and assume that the *sampling set consists of a finite sequence of independent random variables.* Instead of worst case estimates, i.e., inequalities within mathematical analysis, we will seek probabilistic estimates (from the realm of probability theory). With this underlying philosophy, we will pursue the following objectives:

(a) We seek to explain and predict the performance of the existing numerical algorithms.

(b) We estimate the distribution of the condition numbers of the associated Vandermonde-like and Toeplitz-like matrices.

(c) We investigate the asymptotic behavior of condition numbers as the number of samples r tends to infinity.

The randomization of the sampling points seems to be a new idea in the investigation of numerical sampling algorithms. So far random sampling has been investigated by Seip-Ulanovskii [31] and Chistyakov-Lyubarskii-Pastur [9, 10] for entire functions of exponential type of one complex variable. These results rely on the deep characterization of deterministic sampling sets [25, 26] and, to our knowledge, cannot be extended to higher dimensions.

By contrast, our main contribution is to sampling theory for functions of several variables. In higher dimensions there is currently no satisfactory deterministic theory, and our analysis provides the first clues that existing algorithms and methods do really work. From a more applied point of view, our results suggest that random sampling of images or higher-dimensional objects may be a successful strategy to capture the essential information of multi-dimensional objects while preserving numerical efficiency and stability.

Description of Results. We now describe the main results.

Let $\mathcal{P}_M$ be the space of trigonometric polynomials on $\mathbb{R}^d$ of degree M and period 1, that is, $\mathcal{P}_M$ consists of all functions on $\mathbb{R}^d$ of the form

$$p(x) = \sum_{k\in[-M,M]\cap\mathbb{Z}^d} a_k e^{2\pi i k\cdot x} . \tag{1}$$

Note that the (distributional) Fourier transform of $p \in \mathcal{P}_M$ is $\hat{p} = \sum_{k\in[-M,M]\cap\mathbb{Z}^d} a_k\delta_k$, so $\operatorname{supp}\hat{p} \subseteq [-M,M]^d$. The parameter M can be interpreted as the "band-width", and indeed trigonometric polynomials have been shown to be the appropriate finite-dimensional model for band-limited functions [19, 20].

Now assume that the samples $p(x_j), j = 1,\dots,r$, of some trigonometric polynomial $p \in \mathcal{P}_M$ are given for some sampling set $\mathcal{X} = \{x_j : j = 1,\dots,r\}$. By our normalization, we may assume that the sampling set $\mathcal{X}$ is contained in the unit cube $[0,1]^d$. Our goal is to reconstruct or to approximate p. Equivalently, we want to determine the coefficients a_k of p from the samples $p(x_j)$. This task can be seen as a non-uniform discrete Fourier transform and is a frequent task in data processing [8, 13, 14, 28, 33].

In its simplest form, the reconstruction of p amounts to solving the r equations

$$\sum_{k\in[-M,M]^d\cap\mathbb{Z}^d} a_k e^{2\pi i k\cdot x_j} = p(x_j) = y_j \qquad j = 1,\dots,r$$

for the coefficient vector $\mathbf{a} = (a_k)_{k\in\mathbb{Z}^d\cap[-M,M]^d}$. This system of equations can be written in matrix form as

$$\mathcal{U}\mathbf{a} = \mathbf{y} , \tag{2}$$

where $\mathcal{U}$ is the matrix with entries $\mathcal{U}_{jk} = e^{2\pi i k\cdot x_j}, k \in \mathbb{Z}^d \cap [-M,M]^d, j = 1,\dots,r$, and $\mathbf{y}$ is the target vector $\mathbf{y} = (y_j)_{j=1,\dots,r}$. Alternatively, one may try to find $\mathbf{a}$ from the normal equations [17]

$$\mathcal{U}^*\mathcal{U}\mathbf{a} = \mathcal{U}^*\mathbf{y} . \tag{3}$$

In this case the matrix $\mathcal{T} = \mathcal{U}^*\mathcal{U}$ has entries

$$\mathcal{T}_{kl} = \sum_{j=1}^{r} e^{-2\pi i(k-l)\cdot x_j} \qquad k,l \in [-M,M]^d \cap \mathbb{Z}^d .$$

The matrices of these linear systems are highly structured, $\mathcal{U}$ is a *Vandermonde-like matrix* , and $\mathcal{T}$ is a positive semi-definite $D\times D$-matrix with a *block-Toeplitz structure*. Both structures have been successfully exploited for fast numerical algorithms [16, 22, 28, 35].

However, before the numerical analysis of the sampling problem can be undertaken, we need to settle a fundamental theoretical issue: Is either of the equations (2) or (3) solvable? Note that both matrices $\mathcal{U}$ and $\mathcal{T}$ depend on the sampling points x_j as parameters. Therefore we ask more precisely *for which sampling set $\mathcal{X}$ does $\mathcal{U}$ have full rank, or equivalently, when is $\mathcal{T}$ invertible?*

In dimension $d = 1$, $\mathcal{T}$ is invertible if and only if $r \geq 2M + 1$ (the number of sampling points is greater than the dimension of the space). In higher dimensions no criterion for the invertibility of $\mathcal{T}$ is known, and useful results are sparse. See [22] for a discussion.

In the spirit of probability theory we model the sampling set as a sequence of independent, identically distributed random variables (i.i.d. RVs) in $[0, 1]^d$. This means that we treat the sampling points as a sequence of functions $x_j = x_j(\omega)$ on some probability space $(\Omega, \mathbb{P})$. Thus the matrices $\mathcal{U}$ and $\mathcal{T}$ are now random matrices, and their determinants, eigenvalues, and singular values are random variables on $(\Omega, \mathbb{P})$ that depend on the sampling set in a rather complicated way.

The first theorem guarantees the generic invertibility of $\mathcal{T}$.

Theorem 1.1. *Assume that the finite sequence of RVs $x_1, \dots, x_r$, satisfies the following properties:*

(a) $r \geq (2M + 1)^d$.

(b) The x_j's are independent.

(c) The distribution μ_j of each x_j is absolutely continuous with respect to Lebesgue measure on $[0, 1]^d$.

Then with probability one the Toeplitz-like matrix $\mathcal{T}$ is invertible.

Estimates for the Condition Number. For a stable numerical solution of either of the systems (2) and (3) we need effective invertibility of $\mathcal{T}$. This is usually measured by the condition number $\kappa(\mathcal{T})$ of $\mathcal{T}$. (The condition number $\kappa(M)$ of a rectangular matrix is the ratio of largest to smallest singular value [17]; for a positive-definite square matrix, this is simply the ratio of the largest to the smallest eigenvalue.)

To estimate the condition numbers of $\mathcal{U}$ and $\mathcal{T}$ we observe that

$$\sum_{j=1}^{r} |p(x_j)|^2 = \langle \mathbf{y}, \mathbf{y} \rangle = \langle \mathcal{U}\mathbf{a}, \mathcal{U}\mathbf{a} \rangle = \langle \mathcal{U}^*\mathcal{U}\mathbf{a}, \mathbf{a} \rangle = \langle \mathcal{T}\mathbf{a}, \mathbf{a} \rangle . \tag{4}$$

Consequently, if we can prove an inequality of the form

$$A\|p\|_2^2 \leq \sum_{j=1}^{r} |p(x_j)|^2 \leq B\|p\|_2^2 \qquad \forall p \in \mathcal{P}_M \tag{5}$$

then the largest (smallest) eigenvalue of $\mathcal{T}$ is at most B (at least A), since $\|p\|_2 = \|\mathbf{a}\|_2$. Consequently, (5) implies the estimates

$$\kappa(\mathcal{T}) \leq \frac{B}{A} \quad \text{and} \quad \kappa(\mathcal{U}) \leq \left(\frac{B}{A}\right)^{1/2} . \tag{6}$$

Our main theorem is the following asymptotic estimate for the condition numbers of $\mathcal{T}$ or $\mathcal{U}$ as $r \to \infty$.

Theorem 1.2. *Assume that $\mathcal{X} = \{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables uniformly distributed over $[0,1]^d$. There exist constants $A, B > 0$ depending only on the band-width M and the dimension d such that for any $\mu \in (0,1)$, the sampling inequality*

$$(1-\mu) r \|p\|_2^2 \le \sum_{j=1}^{r} |p(x_j)|^2 \le (1+\mu) r \|p\|_2^2 \qquad \forall p \in \mathcal{P}_M \tag{7}$$

holds with probability at least

$$1 - A e^{-Br \frac{\mu^2}{1+\mu}}.$$

Consequently with the same probability estimate the Toeplitz-type matrix $\mathcal{T}$ has condition number $\kappa(\mathcal{T}) \le \frac{1+\mu}{1-\mu}$ and the Vandermonde-like matrix $\mathcal{U}$ has condition number $\kappa(\mathcal{U}) \le \sqrt{1+\mu}/\sqrt{1-\mu}$.

For a fixed threshold $\theta > 1$, the probability that $\kappa(\mathcal{T}) \le \theta$ converges to 1 exponentially fast as the number of samples increases. With some poetic license, we may therefore say that *oversampling improves the condition number*.

We will give two proofs of this result. The first proof is by reduction to a deterministic result. We estimate the probability that the conditions of an existing deterministic result from [22] are satisfied. With this approach we obtain explicit estimates for the constants. The second proof uses a version of the powerful metric entropy method, see [4,5,12] for just a few of its applications to probability theory. This approach is genuinely asymptotic and does not yield effective estimates of the constants. The main advantage of this method is its flexibility and generality. To demonstrate the power of this approach we will formulate versions of Theorem 1.2 for ordinary polynomials in several variables, almost periodic functions, and for spherical harmonics on the sphere (Section 6).

As a consequence of Theorem 1.2 we obtain the following law of the iterated logarithm.

Corollary 1.3. *If $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are uniformly distributed over $[0,1]^d$, then*

$$\limsup_{r\to\infty} \frac{\sup_{p\in\mathcal{P}} |\sum_{j=1}^{r} [|p(x_j)|^2 - \|p\|_2^2]|}{\sqrt{r \log\log r}\,\|p\|_2^2} = c, \qquad a.s. \tag{8}$$

for some positive constant c of order $D = (2M+1)^d$.

With less precision, but more intuitively, the corollary says that with probability one, the condition number of the sampling problem is

$$\kappa(\mathcal{T}) \le (r + c\sqrt{r\log\log r})/(r - c\sqrt{r \log\log r}) \approx 1 + 2c\Big(\frac{\log\log r}{r}\Big)^{1/2},$$

whenever r is large enough.

Our main theorems validate existing numerical algorithms for non-uniform sampling sets in higher dimensions. Furthermore, they make precise in which sense random sampling of multidimensional objects is better than deterministic sampling.

The paper is organized as follows. In Section 2 we collect some facts about multivariate trigonometric polynomials and explain the idea of the simplest numerical algorithms. In Section 3 we prove Theorem 1.1 about the almost certain solvability of the sampling problem. In Section 4 we provide the first proof of Theorem 1.2 and show a probabilistic covering result that may be of independent interest. In Section 5 we develop the metric entropy approach and give a second proof of Theorem 1.2 for the asymptotic estimate of the condition number. Furthermore, we develop some consequences of our main theorem. In Section 6 we discuss extensions of the metric entropy method to other sampling problems.

2. Sampling of Trigonometric Polynomials

We first collect the background information on sampling of trigonometric polynomials and some of the numerical aspects that motivated our investigation.

By $\mathcal{X} = \{x_j : j = 1, \dots, r\}$ we denote a sampling set of r (distinct) points in $[0,1]^d$.

The space of trigonometric polynomials on $\mathbb{R}^d$ of degree M and period 1 in each variable is

$$\mathcal{P}_M = \{p : p(x) = \sum_{k\in[-M,M]^d\cap\mathbb{Z}^d} a_k e^{2\pi i k\cdot x}\} . \tag{9}$$

REMARKS: 1. The vector space $\mathcal{P}_M$ has dimension $D = (2M+1)^d$. This implies that we need at least $(2M+1)^d$ data points in order to recover a polynomial $p \in \mathcal{P}_M$.

2. The parameter M can be interpreted as the "bandwidth" and measures the permissible amount of oscillation (smoothness). We will assume that M is given, but note that the determination of the optimal bandwidth is an important step in the practical application of sampling algorithms [36].

3. On $\mathcal{P}_M$ the following estimates between equivalent norms hold:

$$\begin{aligned} \|p\|_2^2 &= \int_{[0,1]^d} |p(x)|^2\, dx = \|\mathbf{a}\|_2 \\ \|p\|_\infty &\le D^{1/2}\,\|\mathbf{a}\|_2 = D^{1/2}\,\|p\|_2 \\ \|p\|_4^4 &\le \|p\|_\infty^2 \|p\|_2^2 \le D\|p\|_2^4 . \end{aligned} \tag{10}$$

The reconstruction of $p \in \mathcal{P}_M$ from given samples $\{p(x_j) : j = 1, \dots, r\}$ amounts to solving the following system of r equations:

$$\sum_{k\in[-M,M]^d\cap\mathbb{Z}^d} a_k e^{2\pi i k\cdot x_j} = f(x_j) = y_j \qquad j = 1, \dots, r . \tag{11}$$

Introducing the matrices $\mathcal{U}$ and $\mathcal{T}$ with entries

$$U_{jk} = e^{2\pi i k\cdot x_j} \qquad j = 1, \dots, r, k \in [-M,M]^d \cap \mathbb{Z}^d \tag{12}$$

$$\mathcal{T}_{kl} = (\mathcal{U}^*\mathcal{U})_{kl} = \sum_{j=1}^r e^{-2\pi i(k-l)\cdot x_j}\; k,l \in [-M,M]^d \cap \mathbb{Z}^d , \tag{13}$$

we can then formulate the sampling problem for $\mathcal{P}_M$ in several distinct ways.

Lemma 2.1. *The following are equivalent:*
(i) The equations (11) *possess a unique solution in* $\mathcal{P}_M$*.*
(ii) The Vandermonde-type matrix has full rank and $r \geq D$*.*
(iii) There exist $A, B > 0$ *such that*

$$A\|\mathbf{a}\|_2 \leq \|\mathcal{U}\mathbf{a}\|_2 \leq B\|\mathbf{a}\|_2 \qquad \forall \mathbf{a} \in \mathbb{C}^D .$$

(iv) The $D \times D$ *Toeplitz-like matrix* $\mathcal{T}$ *is invertible.*
(v) There exist $A, B > 0$ *such that*

$$A\|p\|_2 \leq \sum_{j=1}^{r} |p(x_j)|^2 \leq B\|p\|_2 \qquad \forall p \in \mathcal{P}_M . \tag{14}$$

If any of (i)-(v) hold, we say that $\mathcal{X}$ is a *set of stable sampling* for $\mathcal{P}_M$ [24].

Despite its lack of mathematical substance, this lemma is useful because each of the criteria may be used as a starting point for the theoretical or numerical investigation of the sampling problem. For the mathematical analysis the *sampling inequality* (14) is most appropriate, because it is invites the use the analytic methods. For the numerical solution of the sampling problem, the linear algebra criteria (ii), (iii), and (iv) are most useful, because the theory of structured matrices offers fast solution techniques.

A numerical algorithm for the solution of (11) could then be based on the following steps:

ALGORITHM

Input. Given a sampling set $\mathcal{X} = \{x_j : j = 1, \dots, r\} \subseteq [0,1]^d$ and a data vector $\mathbf{y} = \{y_j : j = 1, \dots, r\}$. Assume that $\mathcal{T}$ defined in (13) is invertible.

Step 1. Compute $\mathbf{b} = \mathcal{U}^*\mathbf{y}$, i.e.,

$$b_k = \sum_{j=1}^{r} e^{-2\pi i k \cdot x_j}\, y_j \qquad \text{for } k \in [-M, M]^d \cap \mathbb{Z}^d . \tag{15}$$

Step 2. Solve the system of equations

$$\mathbf{a} = \mathcal{T}^{-1}\mathbf{b} . \tag{16}$$

Step 3. Compute $p \in \mathcal{P}_M$ by

$$p(x) = \sum_{k \in I_M} a_k e^{2\pi i k \cdot x} . \tag{17}$$

Then p is the (unique) least square approximation of the given data vector $\mathbf{y}$ in the sense that

$$\sum_{j=1}^{r} |y_j - p(x_j)|^2 = \min_{q \in \mathcal{P}_M} \sum_{j=1}^{r} |y_j - q(x_j)|^2 . \tag{18}$$

If $\mathbf{y}$ arises as the sampled vector of a polynomial $p \in \mathcal{P}_M$, i.e., $y_j = p(x_j)$, then this algorithm provides the exact reconstruction of p.

REMARKS: 1. The numerical implementation of this idea is often referred to as the ACT-algorithm. The decisive step is the solution of matrix equation $\mathcal{T}\mathbf{a} = \mathbf{b}$ in Step 2. Since $\mathcal{T}$ is a positive-definite Toeplitz-like matrix, the exploitation of this structure in conjunction with block Toeplitz solvers and conjugate gradient algorithms have led to fast and efficient reconstruction algorithms in higher dimensions [29, 35]. For numerical issues and real applications we refer to [22].

2. Since the condition numbers of $\mathcal{U}$ and $\mathcal{T}$ are related by $\kappa(\mathcal{T}) = \kappa(\mathcal{U})^2$, it may be better to solve the Vandermonde-type system $\mathcal{U}\mathbf{a} = \mathbf{y}$ directly; see the work of Potts and Steidl [27].

3. Invertibility Almost Surely

We first establish that the reconstruction algorithm discussed in Section 2 works almost surely. In dimension $d = 1$, $\mathcal{T}$ is invertible if and only if $r \geq 2M + 1$. In higher dimensions, a complete and effective characterization of the invertibility seems out of reach. For this reason we use a probabilistic approach.

First a lemma in which λ will denote Lebesgue measure.

Lemma 3.1. *Let $p \in \mathcal{P}_M$ be a trigonometric polynomial in d variables. Then its zero set $\mathcal{Z}(p) = \{x \in [0,1]^d : p(x) = 0\}$ has Lebesgue measure 0.*

Proof. This fact is well known; we provide its easy proof for the sake of completeness.

Fix $x_1, \dots, x_{d-1} \in [0,1]^d$; then $P(x_1, \dots, x_{d-1}, x_d)$ is a trigonometric polynomial in one variable x_d of degree M and has thus at most $2M + 1$ zeros. The set $\{x \in [0,1] : (x_1, \dots, x_{d-1}, x) \in \mathcal{Z}(p)\}$ has Lebesgue measure 0. This is true for every choice of $x_1, \dots, x_{d-1}$, so by Fubini's Theorem, we obtain that

$$\lambda(\mathcal{Z}(p)) = \int_{[0,1]^{d-1}} \Big(\int_{[0,1]} \chi_{\mathcal{Z}(p)}(x_1, \dots, x_{d-1}, x)\, dx \Big)\, dx_1 \dots dx_{d-1} = 0\,,$$

as desired. ∎

The following result is a first indication why in practice no serious problems have occurred in the application of multidimensional sampling algorithms.

Theorem 3.2. *Assume that the random variables $\{x_1, \dots, x_r\}$ are independent and that the distribution μ_j of each x_j is absolutely continuous with respect to Lebesgue measure on $[0,1]^d$.*

Then the Vandermonde-like matrix $\mathcal{U}$ is of full rank almost surely. If, in addition, $r \geq D = (2M+1)^d$, then the Toeplitz-like matrix $\mathcal{T} = \mathcal{U}^*\mathcal{U}$ is invertible almost surely.

Proof. Let $m_1, \dots, m_D$ be an enumeration of the index set $[-M, M] \cap \mathbb{Z}^d$ over which we are summing, and let C_N be the $N \times N$-matrix with entries

$$C_{\ell j} = e^{i m_\ell \cdot x_j} \qquad 1 \leq \ell, j \leq N\,.$$

Then C_N depends on the sampling points $x_1, \dots x_N$, and we may define the "bad" set

$$\mathcal{B}_N = \{(x_1, \dots, x_N) \in ([0,1]^d)^N : \det C_N = 0\} .$$

We claim that $\lambda(\mathcal{B}_N) = 0$ for all $N \le \min(r, D)$, and prove this by induction over N. This is certainly true for $N = 1$. So assume that $N < \min(r, D)$ and that $(x_1, \dots, x_N) \notin \mathcal{B}_N$.

Let $a_\ell = (C_{\ell,1}, \dots, C_{\ell,N})$, $\ell \le N$, be the ℓ-th row of C_N and let $a_{N+1} = (C_{N+1,1}, \dots, C_{N+1,N})$. Since C_N is invertible, there exist coefficients $b_\ell = b_\ell(x_1, \dots x_N) \in \mathbb{C}$, not all 0, such that

$$a_{N+1} = b_1 a_1 + \cdots b_N a_N .$$

By looking at the $(N+1)$-st column of C_{N+1}, we find that C_{N+1} is invertible if and only if $C_{N+1,N+1} \neq b_1 C_{1,N+1} + \cdots + b_N C_{N,N+1}$, or if and only if

$$e^{i m_{N+1} \cdot x_{N+1}} \neq b_1 e^{i m_1 \cdot x_{N+1}} + \cdots + b_N e^{i m_N \cdot x_{N+1}} .$$

In other words, C_{N+1} is invertible if x_{N+1} is NOT in the set

$$D_N = D_N(x_1, \dots, x_N) = \{x \in [0,1]^d : e^{i m_{N+1} \cdot x} = b_1 e^{i m_1 \cdot x} + \cdots + b_N e^{i m_N \cdot x}\} .$$

For fixed $(x_1, \dots, x_N) \in ([0,1]^d)^N$, D_N is the zero set of some trigonometric polynomial and by Lemma 3.1 D_N has Lebesgue measure 0 in $[0,1]^d$.

Since the bad set $\mathcal{B}_{N+1}$ is contained in $\{(x_1, \dots, x_N, x_{N+1}) \in ([0,1]^d)^{N+1} : x_{N+1} \in D_N(x_1, \dots, x_N)\}$, we see by Fubini's Theorem that

$$\begin{aligned} \lambda(\mathcal{B}_{N+1}) &= \int_{([0,1]^d)^N} \Big(\int_{[0,1]^d} \chi_{B_{N+1}}(x_1, \dots, x_N, x_{N+1})\, dx_{N+1} \Big)\, dx_1 \dots dx_N \\ &\le \int_{([0,1]^d)^N} \int_{[0,1]^d} \lambda(D_N(x_1, \dots, x_N)) dx_1 \dots dx_N = 0 . \end{aligned}$$

The induction step is proved.

If $r \le D$, then C_r is invertible for almost every choice of $x_1, \dots, x_D$, where "almost every" is with respect to Lebesgue measure λ. Consequently, the $r \times D$ matrix $\mathcal{U}$ has full rank. If $r \ge D$, this also implies that the $D \times D$ square matrix $\mathcal{T} = \mathcal{U}^* \mathcal{U}$ is invertible for almost every choice of $x_1, \dots, x_D$.

Since the distribution μ_j of x_j is absolutely continuous with respect to λ, the bad set $\mathcal{B}_D$ also has measure 0 with respect to $\mu_1 \times \cdots \times \mu_D$. ∎

Corollary 3.3. *The Toeplitz-like matrix $\mathcal{T}$ is invertible under each of the following hypotheses on the sampling set:*

(a) The $x_j, j = 1, \dots, r$, are i.i.d. random variables, each of which is uniformly distributed over $[0,1]^d$.

(b) The sampling set is a random perturbation of a uniform sampling set, i.e., it is some enumeration of $\{\frac{1}{N} k + \delta_k : k \in \mathbb{Z}^d \cap [0, N-1]^d\}$ where $N \ge 2M + 1$ and the δ_k are i.i.d. random variables uniformly distributed over a neighborhood of 0.

4. A Covering Results and Reduction to Deterministic Estimates

Theorem 1.1 guarantees that an implementation of Algorithm 2 will work in principle. However, numerical invertibility requires a reasonable bound on the condition number of $\mathcal{T}$ or of $\mathcal{U}$.

This is already a serious problem in dimension $d = 1$. It is easy to construct sampling sets in $[0,1]$ for which the corresponding Toeplitz matrix has condition number of the order 10^{15} [16]. While such a matrix is invertible in theory, for practical purposes it may be considered to be non-invertible.

As a next step we therefore turn to estimates for the condition number of the block Toeplitz matrix $\mathcal{T}$. For this we combine a deterministic result with a probabilistic statement on coverings.

We work with the metric $d(x,y) = \min_{k\in\mathbb{Z}^d} \|x-y+k\|_\infty$ on the torus $\mathbb{T}^d \sim [0,1)^d$ and the associated cubes of side-length 2ρ

$$B(x,\rho) = \{y \in [0,1]^d : d(y,x) \le \rho\} = x + [-\rho,\rho]^d \,.$$

To every sequence of sampling points $\{x_j : j \in \mathbb{N}\} \subseteq [0,1]^d$, let $\{V_j\}$ we assign the "distance function"

$$\delta(r) = \inf\{s : \bigcup_{j=1}^{r} B(x_i, s) \supset [0,1]^d\} \,. \tag{19}$$

The quantity $2\delta(r)$ can be interpreted as the maximum distance of any of the first r sampling points x_j to its next neighbor. Let $V_j, j = 1,\dots,r$, be Voronoi regions

$$V_j = \{y \in [0,1]^d : d(y,x_j) \le d(y,x_k), k \ne j, 1 \le j,k \le r\}$$

and $w_j = \lambda(V_j)$ and consider the weighted Toeplitz-like matrix $\mathcal{T}^w$ with entries

$$\mathcal{T}^w_{kl} = (\mathcal{U}^*\mathcal{U})_{kl} = \sum_{j=1}^{r} w_j e^{-2\pi i(k-l)\cdot x_j} \qquad k,l \in [-M,M]^d \cap \mathbb{Z}^d \,.$$

Then it is possible to show the following deterministic theorem [18, 22].

Theorem 4.1. *If*

$$\delta(r) < \frac{\log 2}{2\pi M d} \,, \tag{20}$$

then, for all $p \in \mathcal{P}_M$,

$$(2 - e^{2\pi d M\delta})^2 \|p\|_2^2 \le \sum_{j=1}^{r} |p(x_j)|^2 w_j = \langle a, \mathcal{T}^w a\rangle \le 4\|p\|_2^2 \,. \tag{21}$$

Consequently, the condition number of $\mathcal{T}^w$ can be estimated by

$$\kappa(\mathcal{T}^w) \le \frac{4}{(2 - e^{2\pi d M\delta})^2} \,, \tag{22}$$

and both $\mathcal{T}$ and $\mathcal{T}^w$ are invertible.

REMARKS: 1. The specific choice of weights w_j is crucial for the explicit estimate (22). In the numerical implementation of the algorithm of Section 2, they serve as a simple and cheap preconditioner.

2. In higher dimensions (22) is far from being optimal, since it depends on the dimension d. It is an open problem to obtain improvements to this estimate. For a related result for band-limited functions, see [7].

We next suppose that the sampling points form an infinite sequence of i.i.d. independent random variables $x_j, j \in \mathbb{N}$. We first investigate how the distribution of the associated sequence of random variables $\delta(r)$ depends on the number of sampling points r.

Theorem 4.2. *If $X = \{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables uniformly distributed over $[0,1]^d$, then for every $r, N \in \mathbb{N}$*

$$\mathbb{P}\Big(\delta(r) > 1/N\Big) \le N^d(1-N^{-d})^r \le N^d e^{-r/N^d}\,. \tag{23}$$

Consequently, $\kappa(\mathcal{T}^w) \le 4(2 - e^{2\pi Md/N})^{-2}$ and both $\mathcal{T}^w$ and $\mathcal{T}$ are invertible with probability at least

$$1 - N^d(1-N^{-d})^r \ge 1 - N^d e^{-r/N^d}.$$

Proof. Divide $[0,1]^d$ into N^d disjoint subcubes of side length $1/N$, i.e., $[0,1]^d = \bigcup_{j=1}^r B(c_j, \frac{1}{2N})$, where the c_j are the centers of these subcubes. Note that if a subcube contains a point x_j, then that subcube is contained in $B(x_j, 1/N)$. So if each of these subcubes contains at least one of the x_j, we conclude $\delta(r) \le 1/N$.

Since the $x_j, j = 1, \dots, r$, are chosen independently and uniformly, the number of x_j's in any cube is a binomial random variable. Thus the probability that a particular subcube is empty is

$$(1 - N^{-d})^r\,,$$

(since N^{-d} is the probability that any particular x_j is in this subcube and there are r points). Since there are N^d subcubes altogether, the probability that at least one of the subcubes is empty is bounded by

$$N^d(1-N^{-d})^r\,. \tag{24}$$

If $\delta(r) > 1/N$, then at least one of the subcubes must be empty, which proves the left-hand inequality of (23). The right-hand side follows from the obvious inequality $(1-N^{-d})^r = e^{r\log(1-N^{-d})} \le e^{-r/N^d}$.

The estimate for the condition number of $\mathcal{T}^w$ and the invertibility of $\mathcal{T}$ now follow from Theorem 4.1. ∎

REMARK: For (20) we need that $\frac{1}{N} < \frac{\log 2}{2\pi Md}$; this means that we need at least $r = N^d \ge \left(\frac{2\pi Md}{\log 2}\right)^d \approx \left(\frac{\pi d}{\log 2}\right)^d D$ sampling points before Theorem 4.2 becomes effective.

Next we derive an asymptotic result for $\delta(r)$, which may be of independent interest.

Theorem 4.3. *Assume that* $\{x_j : j \in \mathbb{N}\}$ *is a sequence of i.i.d points uniformly distributed in* $[0,1]^d$. *Then*

$$\limsup_{r\to\infty} \frac{\delta(r)}{(\log r/r)^{1/d}} = c\,, \qquad a.s. \tag{25}$$

for some constant $c \in [\frac{1}{4}, 2^{1+1/d}]$.

Thus for r sampling points the maximum distance to the nearest neighbor is roughly $(\log r/r)^{1/d}$. For comparison, for the $r = N^d$ equispaced points $\{\frac{k}{N} : k \in [0,N] \cap \mathbb{Z}^d\}$, we have $\delta(r) = \frac{1}{2N} = \frac{1}{2} r^{-1/d}$. For r randomly distributed points we need an additional logarithmic term.

Proof of Theorem 4.3. **Step 1.** We first show that

$$\limsup_{r\to\infty} \frac{\delta(r)}{(\log r/r)^{1/d}} \le 2^{1+1/d}, \qquad \text{a.s.} \tag{26}$$

Choose $r_k = 2^k$ as the number of points, and let N_k be the greatest integer less than $\left(\frac{r_k}{2\log r_k}\right)^{1/d}$. We divide $[0,1]^d$ into N_k^d disjoint subcubes of side length N_k^{-1}. Let A_k be the event that at least one of the subcubes contains none of the $x_j, j = 1, \dots, r_k$. By (24) we have

$$\mathbb{P}(A_k) \le N_k^d e^{-r_k/N_k^d} \le \frac{r_k}{2\log r_k} e^{-2\log r_k} = \frac{1}{2r_k \log r_k} = \frac{1}{2^{k+1} k \log 2}\,. \tag{27}$$

Therefore $\sum_{k=1}^\infty \mathbb{P}(A_k) < \infty$, and so the Borel-Cantelli lemma [11] implies that the probability of A_k infinitely often is 0. This means for almost every $\omega \in \Omega$ there is a k_0 depending on ω such that for $k \ge k_0$, each of the subcubes of side length N_k^{-1} will contain at least one of the points of $x_1, \dots, x_{r_k}$.

Now for r arbitrary and sufficiently large (depending on ω), choose k such that $r_k \le r < r_{k+1}$. Then each of the subcubes of side length N_k^{-1} will contain at least one of the points $x_1, \dots, x_{r_k}$, hence at least of the points $x_1, \dots, x_r$. Consequently

$$\delta(r) \le \frac{1}{N_k}\,,$$

and thus

$$\left(\frac{r}{\log r}\right)^{1/d} \delta(r) \le \left(\frac{r_{k+1}}{\log r_{k+1}}\right)^{1/d} \delta(r) \le 2^{1/d}(2N_k + 1)\delta(r) \le 2^{1/d}(2 + \frac{1}{N_k})\,.$$

Taking $r \to \infty$ proves (26).

We prove the converse inequality

$$\limsup_{r\to\infty} \frac{\delta(r)}{(\log r/r)^{1/d}} \ge \frac{1}{4}, \qquad \text{a.s.} \tag{28}$$

in several steps.

Step 2. Assume for the moment that we have already chosen a sequence r_k (number of sampling points) and N_k. Then we divide $[0,1]^d$ into N_k^d subcubes of side length N_k^{-1}, and we enumerate the cubes as $C_1, C_2, \dots, C_{N_k^d}$. Let D_j be the event that the cube C_j does not contain any of the points $x_{r_{k-1}+1}, \dots, x_{r_k}$. As in (24) the probability of D_j is given by

$$\mathbb{P}(D_j) = (1 - N_k^{-d})^{r_k - r_{k-1}} . \tag{29}$$

For $j \neq k$, $D_j \cap D_k$ is the event that the region $C_j \cup C_k$ does not contain any of the points $x_{r_{k-1}+1}, \dots, x_{r_k}$. Therefore as in (24) we obtain that

$$\begin{aligned} \mathbb{P}(D_j \cap D_k) &= (1 - 2N_k^{-d})^{r_k - r_{k-1}} \leq \\ &\leq (1 - N_k^{-d})^{2(r_k - r_{k-1})} = \mathbb{P}(D_j)\mathbb{P}(D_k) , \end{aligned} \tag{30}$$

since $1 - 2x \leq (1-x)^2$ for $x \in [0,1]$.

Step 3. Now let B_k be the event that at least one of the first N_k (out of a total of N_k^d) cubes $C_1, \dots, C_{N_k}$ does not contain any of the points $x_{r_{k-1}}, \dots, x_{r_k}$. (In dimension $d = 1$ we take the first N_k^α of N_k cubes for some $\alpha, 1/2 < \alpha < 1 - 1/e$ and modify the following argument slightly.) If we define the random variable Y_k by

$$Y_k = \sum_{j=1}^{N_k} 1_{D_j} ,$$

then $B_k = \{Y_k > 0\}$. To find a lower estimate for the probability of B_k, we use an argument due to Kochen-Stone [23]. Using Cauchy-Schwarz we find that

$$\begin{aligned} \mathbb{E}\, Y_k &= \sum_{l=1}^{N_k} l\, \mathbb{P}(Y_k = l) \\ &\leq \Big(\sum_{l=1}^{N_k} l^2 \mathbb{P}(Y_k = l) \Big)^{1/2} \Big(\sum_{l=1}^{N_k} \mathbb{P}(Y_k = l) \Big)^{1/2} \\ &= (\mathbb{E}\, Y_k^2)^{1/2} (\mathbb{P}(Y_k > 0))^{1/2} , \end{aligned}$$

whence

$$\mathbb{P}(B_k) = \mathbb{P}(Y_k > 0) \geq \frac{(\mathbb{E}\, Y_k)^2}{\mathbb{E}\, Y_k^2} . \tag{31}$$

On the other hand,

$$\mathbb{E}\, Y_k = \sum_{j=1}^{N_k} \mathbb{P}(D_j) = N_k \mathbb{P}(D_j)$$

and by (30)

$$\begin{aligned} \mathbb{E}\, Y_k^2 &= \sum_{j=1}^{N_k} \mathbb{P}(D_j) + \sum_{k \neq j} \mathbb{P}(D_j \cap D_k) \\ &\leq \mathbb{E}\, Y_k + \sum_{k \neq j} \mathbb{P}(D_j)\mathbb{P}(D_k) \\ &\leq \mathbb{E}\, Y_k + (\mathbb{E}\, Y_k)^2. \end{aligned}$$

Substituting into (31), we obtain

$$\mathbb{P}(B_k) \geq \frac{\mathbb{E}\, Y_k}{1+\mathbb{E}\, Y_k}\,. \tag{32}$$

Step 4. Finally we choose $r_k = e^{e^k}$ and N_k the least integer $\geq (2^d r_k/\log r_k)^{1/d}$. Then

$$\mathbb{P}(D_j) = (1-N_k^{-d})^{r_k - r_{k-1}} \geq \left(1 - \frac{\log r_k}{2^d r_k}\right)^{r_k}\,.$$

Since $\lim_{x\to\infty} x^{1/2^d}\big(1-\frac{\log x}{2^d x}\big)^x = 1$, we have $\big(1-\frac{\log x}{2^d x}\big)^x \geq \frac{1}{2}x^{-1/2^d}$ for x sufficiently large, and consequently

$$\mathbb{E}\, Y_k = N_k \mathbb{P}(D_j) \geq \left(\frac{2^d r_k}{\log r_k}\right)^{1/d} \frac{1}{2} r_k^{-1/2^d} = r_k^{\frac{1}{d}-\frac{1}{2^d}}/\log r_k \geq 1 \tag{33}$$

for sufficiently large k ($k \geq 3$). Now (32) implies that $\mathbb{P}(B_k) \geq 1/2$ and so $\sum_{k=1}^\infty \mathbb{P}(B_k) = \infty$. Finally we observe that the events B_k are independent, because they depend on disjoint segments of the sequence $x_j, j \in \mathbb{N}$. Therefore the second part of the Borel-Cantelli lemma [11] implies that the probability of B_k infinitely often is 1. This means that for almost every ω there is an infinite subsequence of k's (depending on ω) such that $\omega \in B_k$.

Step 5. It remains to consider the event E_k that one of the points $x_1, \dots, x_{r_{k-1}}$ is in $\bigcup_{j=1}^{N_k} C_j$. Since the volume of $\bigcup_{j=1}^{N_k} C_j$ is $N_k \cdot N_k^{-d}$, the probability that a particular x_j is in this set is N_k^{1-d}. There are r_{k-1} points to consider, so as in (24)

$$\mathbb{P}(E_k) \leq r_{k-1} N_k^{1-d}$$

By our choices of r_k and N_k, we have $\sum_{k=1}^\infty \mathbb{P}(E_k) < \infty$, and so by the Borel-Cantelli lemma once again, the probability of E_k infinitely often is 0.

Combining Steps 4 and 5 we conclude that with probability 1, infinitely often at least one of the C_ℓ with $\ell \leq N_k$ will contain none of the points $x_1, \dots, x_{r_k}$. Since C_ℓ contains none of these x_j, the center of C_ℓ is not contained in $\bigcup_{j=1}^{r_k} B(x_j, 1/(2N_k))$. Consequently $\delta(r_k) > 1/(2N_k)$ for infinitely many k almost surely. So

$$\delta(r_k)\left(\frac{r_k}{\log r_k}\right)^{1/d} \geq \delta(r_k) N_k/2 \geq 1/4$$

and (28) is proved.

Step 6. It is clear that if we omit the first M points $x_1, \dots, x_M$ for any fixed integer M, then this will not affect the value of $\limsup \delta(r)/(\log r/r)^{1/d}$. Therefore this random variable is measurable with respect to the tail σ-field of the sequence $x_1, x_2, \dots$ By the Kolmogorov's zero-one law, the value of this random variable must be constant almost surely [11, p. 254]. This completes the proof. ■

5. Asymptotic Estimates of the Condition Number

In the previous section we have combined a deterministic argument with a covering argument. Essentially we have calculated the probability that a random sampling set satisfies the sufficient condition already known from deterministic sampling theory.

In this section we develop an alternative approach that is based on a metric entropy argument such as the ones used in [12]. This approach does not rely on deterministic sampling results and can therefore be adapted to other sampling models. On the other hand, it is difficult to keep track of the constants involved, and thus the results are only efficient for large sampling sets.

Once again we start with an infinite sequence of i.i.d. random variables $\{x_j : j \in \mathbb{N}\}$, each of which is uniformly distributed over $[0,1]^d$. Our goal is to estimate the quantity $\sum_{j=1}^r |p(x_j)|^2 - r\|p\|_2^2$ and its distribution as a function of the number of sampling points r.

For every $p \in \mathcal{P}_M$ we introduce the random variable $Y_j(p) = |p(x_j)|^2 - \|p\|_2^2$. To obtain a sampling inequality of the form $A\|p\|_2^2 \le \sum_{j=1}^r |p(x_j)|^2 \le B\|p\|_2^2$, we have to estimate the probability distribution of the random variable

$$\sup_{p\in\mathcal{P}_M, \|p\|_2=1} \sum_{j=1}^r Y_j(p) \, .$$

This is accomplished in the following theorem.

Theorem 5.1. *If $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are uniformly distributed over $[0,1]^d$, then there exist constants $A, B > 0$ depending on d and M, such that*

$$\mathbb{P}\Big(\sup_{p\in\mathcal{P}_M, \|p\|_2=1} \sup_{s\le r} \Big| \sum_{j=1}^s Y_j(p) \Big| \ge \lambda \Big) \le A \exp\Big(-B \frac{\lambda^2}{r+\lambda} \Big) \tag{34}$$

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

For the distribution of a sum of random variables we use Bernstein's inequality [6] in the following form.

Proposition 5.2. *Let $Y_j, j = 1, \dots, r$, be a sequence of bounded, independent random variables with $\mathbb{E}\, Y_j = 0$, $\mathrm{Var} Y_j = \sigma^2$, and $\|Y_j\|_\infty \le M$ for $j = 1, \dots, r$. Then*

$$\mathbb{P}\Big(\Big| \sum_{j=1}^r Y_j \Big| \ge \lambda \Big) \le 2 \exp\Big(-\frac{\lambda^2}{2r\sigma^2 + \frac{2}{3}M\lambda} \Big) \, . \tag{35}$$

To apply (35) to the $Y_j(p)$, we need several simple estimates. It suffices to work with the unit ball of $\mathcal{P}_M$, which we denote by $\mathcal{P}^0 = \{p \in \mathcal{P}_M : \|p\|_2 \le 1\}$.

Lemma 5.3. *Let $p, q \in \mathcal{P}^0$ and $j \in \mathbb{N}$. Then the following identities and inequalities hold:*

$$\mathbb{E}\, Y_j(p) = 0 \tag{36}$$

$$\mathrm{Var}\, Y_j(p) = \|p\|_4^4 - \|p\|_2^4 \le D - 1\,, \tag{37}$$

$$\mathrm{Var}\,(Y_j(p) - Y_j(q)) \le 8\|p - q\|_\infty^2\,, \tag{38}$$

$$\|Y_j(p)\|_\infty \le \|p\|_\infty^2 - \|p\|_2^2 \le (D-1)\,, \tag{39}$$

$$\|Y_j(p) - Y_j(q)\|_\infty \le 2(D^{1/2} + 1)\|p - q\|_\infty\,. \tag{40}$$

Proof. Since each x_j is uniformly distributed over $[0,1]^d$, we have

$$\mathbb{E}\,(Y_j(p)) = \int_{[0,1]^d} \big(|p(x)|^2 - \|p\|_2^2\big)\, dx = 0$$

and consequently (also using (10))

$$\begin{aligned} \mathrm{Var}\, Y_j(p) &= \mathbb{E}\,[Y_j(p)^2] = \int_{[0,1]^d} \big(|p(x)|^2 - \|p\|_2^2\big)^2\, dx \\ &= \|p\|_4^4 - \|p\|_2^4 \le D - 1\,. \end{aligned}$$

since $\|p\|_2 = 1$. Similarly, we obtain

$$\|Y_j(p)\|_\infty = \sup_{\omega\in\Omega} \Big|\, |p(x_j(\omega))|^2 - \|p\|_2^2 \Big| \le \Big| \|p\|_\infty^2 - \|p\|_2^2 \Big| \le D - 1$$

Next, since $\mathbb{E}\, Y_j(p) = 0$, we obtain

$$\begin{aligned} \mathrm{Var}\Big(Y_j(p) - Y_j(q)\Big) &= \mathbb{E}\Big((Y_j(p) - Y_j(q))^2\Big) \\ &= \int_{[0,1]^d} \big(|p(x)|^2 - |q(x)|^2\big)^2\, dx - \Big(\|p\|_2^2 - \|q\|_2^2\Big)^2 \\ &\le \|p - q\|_\infty^2\, \|\, |p| + |q|\, \|_2^2 + \|p - q\|_2^2\big(\|p\|_2^2 + \|q\|_2^2\big) \\ &\le 8\|p - q\|_\infty^2 \end{aligned}$$

The last estimate follows similarly from

$$\begin{aligned} \|Y_j(p) - Y_j(q)\|_\infty &\le \sup_{\omega\in\Omega} \Big|\, |p(x_j(\omega))|^2 - |q(x_j(\omega))|^2 \Big| + \Big| \|q\|_2^2 - \|p\|_2^2 \Big| \\ &\le \|p - q\|_\infty\big(\|p\|_\infty + \|q\|_\infty\big) \;+\; \|q - p\|_2\big(\|p\|_2 + \|q\|_2\big) \\ &\le \|p - q\|_\infty\, D^{1/2}\big(\|p\|_2 + \|q\|_2\big) \;+\; \|q - p\|_\infty\big(\|p\|_2 + \|q\|_2\big) \\ &= 2(D^{1/2} + 1)\, \|p - q\|_\infty \end{aligned}$$

■

Proof of Theorem 5.1. **Step 1: A Metric Entropy Argument.** For a given $\delta > 0$, we construct a δ-net for $\mathcal{P}^0$ with respect to the L^∞-norm as follows. Given $p \in \mathcal{P}^0$ with coefficients $\mathbf{a} = (a_k)_{k\in\mathbb{Z}^d\cap[-M,M]^d}$ and $\|\mathbf{a}\|_2 \le 1$, we approximate the real and imaginary part of each a_k by a number $\frac{\delta}{\sqrt{2}D}\ell, \ell \in \mathbb{Z}$; in other words, we choose a vector $\mathbf{b}$ of the form $\mathbf{b} = \frac{\delta}{\sqrt{2D}}(\ell + im), \ell, m \in \mathbb{Z}^d$, to approximate $\mathbf{a}$. Then for each coordinate $a_k, k \in [-M, M]^d \cap \mathbb{Z}^d$, we have

$$|a_k - b_k| \le \frac{\delta}{D},$$

and so

$$\|\mathbf{a} - \mathbf{b}\|_2 \le \big(D \max_k |a_k - b_k|^2\big)^{1/2} = \frac{\delta}{\sqrt{D}}.$$

Setting $q(x) = \sum_{k \in I_M} b_k e^{2\pi i k \cdot x}$, we obtain

$$\|p - q\|_\infty \le D^{1/2} \|p - q\|_2 \le D^{1/2} \|\mathbf{a} - \mathbf{b}\|_2 = \delta.$$

We denote the δ-net of all $q \in \mathcal{P}^0$ with coefficients of the form $\mathbf{b} = \frac{\delta}{\sqrt{2D}}(\ell + im), \ell, m \in \mathbb{Z}^d, \|\mathbf{b}\|_2 \le 1$, by $\mathcal{A}(\delta)$. The cardinality of $\mathcal{A}(\delta)$ is estimated as follows:

$$\begin{aligned} \operatorname{card} \mathcal{A}(\delta) &= \operatorname{card} \{\mathbf{b} = \frac{\delta}{\sqrt{2D}}(\ell + im), \ell, m \in \mathbb{Z}^D, \|\mathbf{b}\|_2 \le 1\} \\ &= \operatorname{card} \{k \in \mathbb{Z}^{2D} : \|k\|_2 \le \frac{\sqrt{2}D}{\delta}\} \\ &\le c_1 \delta^{-2D}. \end{aligned}$$

where the constant $c_1 \approx \frac{(2\pi)^D}{D} D^{2D}$ is roughly the number of integer lattice points in a ball of radius $\sqrt{2}D$ in $\mathbb{R}^{2D}$.

Given $p \in \mathcal{P}^0$, let p_j be the polynomial in $\mathcal{A}(2^{-j})$ that is closest to p in L^∞ norm, with some convention for breaking ties. Since $\|p - p_j\|_2 \to 0$, we can write

$$Y_j(p) = Y_j(p_0) + (Y_j(p_1) - Y_j(p_0)) + (Y_j(p_2) - Y_j(p_1)) + \cdots.$$

If $\sup_{p \in \mathcal{P}} \sup_{s \le r} |\sum_{j=1}^s Y_j(p)| \ge \lambda$, then either

(a) $\sup_{s \le r} |\sum_{j=1}^s Y_j(p)| \ge \lambda/2$ for some $p \in \mathcal{A}(1)$; or

(b) for some $\ell \ge 1$ and some $p \in \mathcal{A}(2^{-\ell})$ and some $q \in \mathcal{A}(2^{-\ell+1})$ with $\|p - q\|_\infty \le 3 \cdot 2^{-\ell}$ we have $\sup_{s \le r} |\sum_{j=1}^s \big(Y_j(p) - Y_j(q)\big)| \ge \lambda/2(\ell+1)^2$.

(Possibly both (a) and (b) hold.)

If this were not the case, then

$$\begin{aligned} \sup_{s \le r} \Big|\sum_{j=1}^s Y_j(p)\Big| &\le \sup_{s \le r} \Big|\sum_{j=1}^s Y_j(p_0)\Big| + \sup_{s \le r} \sum_{\ell=1}^\infty \Big|\sum_{j=1}^s \Big(Y_j(p_\ell) - Y_j(p_{\ell-1})\Big)\Big| \\ &\le \sum_{\ell=1}^\infty \frac{\lambda}{2\ell^2} = \frac{\pi^2}{12}\lambda < \lambda. \end{aligned}$$

So far the construction is purely deterministic. Now we estimate the probability of each of the events in (a) and (b).

Step 2. For fixed $p \in \mathcal{A}(1)$, the probability of the event in (a) is bounded, using Bernstein's inequality (35) and Lemma 5.3, by

$$\begin{aligned} & 2\exp\Big(-\frac{\lambda^2}{2r\operatorname{Var} Y_j(p) + \frac{2}{3}\lambda\|Y_j(p)\|_\infty}\Big) \\ \le\ & 2\exp\Big(-\frac{\lambda^2}{2r(D-1) + \frac{2}{3}(D-1)\lambda}\Big)\,. \end{aligned}$$

There are at most c_1 polynomials in $\mathcal{A}(1)$, so the probability of (a) is bounded by

$$2c_1\exp\Big(-\frac{\lambda^2}{(D-1)(2r+\frac{2}{3}\lambda)}\Big)\,. \tag{41}$$

Step 3. We estimate (b) in a similar fashion using Lemma 5.3, (38) and (40). If $p \in \mathcal{A}(2^{-\ell})$ and $q \in \mathcal{A}(2^{-\ell+1})$ with $\|p-q\|_\infty \le 3\cdot 2^{-\ell}$, we have

$$\begin{aligned} \mathbb{P}\Big(\sup_{s\le r}|\sum_{j=1}^{s}(Y_j(p)-Y_j(q))| \ &>\ \frac{\lambda}{2(\ell+1)^2}\Big) \\ &\le\ 2\exp\Big(-\frac{\lambda^2/4(\ell+1)^4}{144r2^{-2\ell} + 4\cdot 2^{-\ell}D^{1/2}\lambda/(\ell+1)^2)}\Big) \\ &\le\ 2\exp\Big(-2^{\ell}\,\frac{\lambda^2}{c_3(r(\ell+1)^4 2^{-\ell} + D^{1/2}\lambda(\ell+1)^2)}\Big)\,. \end{aligned}$$

There are $c_1 2^{(2\ell-2)D}$ trigonometric polynomials in $\mathcal{A}(2^{-\ell+1})$, and for each q the number of trigonometric polynomials $p \in \mathcal{A}(2^{-\ell})$ satisfying $\|p-q\|_\infty \le 3\cdot 2^{-\ell}$ is bounded by a constant c_2 independent of q and j (Similar to the count in Step 1, $c_2 \approx \frac{(6\pi)^D}{D}D^{2D}$ is roughly the number of integer lattice points in a ball of radius $3\sqrt{2}D$ in $\mathbb{R}^{2D}$). Finally, this can happen for any ℓ. So the probability in (b) is bounded by

$$\sum_{\ell=1}^{\infty} 2c_1c_2 2^{(2\ell-2)D}\exp\Big(-2^{\ell}\,\frac{\lambda^2}{c_3(r(\ell+1)^4 2^{-\ell} + D^{1/2}\lambda(\ell+1)^2)}\Big)\,. \tag{42}$$

Step 4. Estimate of the sum (42).

Since $(\ell+1)^4 2^{-\ell}$ is bounded above and $2^{\ell/2}/(\ell+1)^2$ is bounded below, the above sum is bounded by

$$\sum_{\ell=1}^{\infty} c_4\exp\Big(-2^{\ell/2}\frac{\lambda^2}{c_5(r+\lambda)} + (2\ell-2)D\log 2\Big) = (\star)\,. \tag{43}$$

We distinguish two cases. Either

$$\frac{\lambda^2}{c_5(r+\lambda)} \ge 64D\,, \tag{44}$$

then

$$2^{\ell/2}\frac{\lambda^2}{c_5(r+\lambda)} \ge 2(2\ell-2)D\log 2, \qquad \text{for all } \ell \ge 1\,,$$

and so

$$(\star) \le \sum_{\ell=1}^{\infty} c_4 \exp\Big(-2^{\ell/2}\frac{\lambda^2}{2c_5(r+\lambda)}\Big).$$

Now we use the fact that $\sum_{\ell=1}^{\infty} e^{-a^\ell x} \le c_6 e^{-x}$ for any $a > 1$ and $x \ge 1$ (with c_6 depending only on a). Consequently the sum in (43) is bounded by

$$(\star) \le c_7 \exp\Big(-\frac{\lambda^2}{c_8(r+\lambda)}\Big).$$

In the second case, (44) does not hold. But then the probability of the event in (b) is at most 1 which is certainly less than or equal to

$$e^{64D} \exp\Big(-\frac{\lambda^2}{c_8(r+\lambda)}\Big).$$

In either case, we have that the probability of the event in (b) is bounded by

$$c_9 \exp\Big(-\frac{\lambda^2}{c_8(r+\lambda)}\Big).$$

Step 5. The statement now follows by combining the bounds for (a) and (b), and so we have

$$\mathbb{P}\Big(\sup_{p\in\mathcal{P}}\sup_{s\le r}\Big|\sum_{j=1}^{s} Y_j(p)\Big| \ge \lambda\Big) \le A\exp\Big(-B\frac{\lambda^2}{r+\lambda}\Big). \tag{45}$$

■

Corollary 5.4. *If $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are uniformly distributed over $[0,1]^d$ and $0 < \mu < 1$, then the sampling inequality*

$$(1-\mu)r\|p\|_2^2 \le \sum_{j=1}^{r} |p(x_j)|^2 \le (1+\mu)r\|p\|_2^2 \qquad \forall p \in \mathcal{P}_M \tag{46}$$

holds with probability at least

$$1 - Ae^{-Br\frac{\mu^2}{1+\mu}}.$$

Consequently with the same probability estimate the Toeplitz-type matrix $\mathcal{T}$ has condition number $\kappa(\mathcal{T}) \le \frac{1+\mu}{1-\mu}$ and also $\kappa(\mathcal{U}) \le \Big(\frac{1+\mu}{1-\mu}\Big)^{1/2}$

Proof. Choose $\lambda = r\mu$ in Theorem 5.1 and observe that the inequality

$$\Big|\sum_{j=1}^{r} |p(x_j)|^2 - r\Big| \le r\mu$$

for all $p \in \mathcal{P}^0$ is equivalent to the sampling inequality (46) for all $p \in \mathcal{P}_M$. ■

From Theorem 5.1 it is straightforward to obtain a law of the iterated logarithm.

Corollary 5.5. *If $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are uniformly distributed over $[0,1]^d$, then*

$$\limsup_{r\to\infty} \frac{\sup_{p\in\mathcal{P}} |\sum_{j=1}^r [|p(x_j)|^2 - \|p\|_2^2]|}{\sqrt{r\log\log r}\,\|p\|_2^2} = c, \qquad a.s. \tag{47}$$

for some constant $c \in [\left(\frac{2}{\pi}\right)^d D - 1, \infty)$.

Proof. Let $r_k = 2^k$ and $\lambda_k = \frac{2}{\sqrt{B}}\sqrt{r_k \log\log r_k}$, where B is the constant from (34). Let

$$C_k = \{\sup_{p\in\mathcal{P}^0} \sup_{s\le r_k} \Big|\sum_{j=1}^s Y_j(p)\Big| > \lambda_k\}.$$

Then for k large enough, we have $r_k > \lambda_k$. So the probability of C_k is bounded by

$$\begin{aligned} \mathbb{P}(C_k) &\le A\exp\Big(-B\frac{\lambda_k^2}{r_k+\lambda_k}\Big) \\ &\le A\exp\Big(-B\frac{\lambda_k^2}{2r_k}\Big) \\ &\le A\exp\Big(-B\frac{4}{B}\frac{r_k\log\log r_k}{2r_k}\Big) \\ &= A\exp\Big(-2\log k\Big) = \frac{A}{k^2}\,. \end{aligned}$$

So $\sum_{k=1}^\infty \mathbb{P}(C_k) < \infty$, and by the Borel-Cantelli lemma, the probability of C_k happening infinitely often is 0.

If $|\sum_{j=1}^r Y_j(p)| > \frac{2}{\sqrt{B}}\sqrt{r\log\log r}$ for some r, we choose k so that $r_{k-1} \le r < r_k$ and observe that C_k holds. (This is the only place where we need the estimate for $\sup_{s\le r} |\sum_{j=1}^s Y_j(p)|$ instead of just $|\sum_{j=1}^r Y_j(p)|$.) So this inequality cannot happen for infinitely many r and we therefore have

$$\limsup_{r\to\infty} \frac{\sup_{p\in\mathcal{P}^0} |\sum_{j=1}^r [|p(x_j)|^2 - r|}{\sqrt{r\log\log r}} \le c', \qquad a.s.$$

for some constant $c' > 0$.

For fixed $p \in \mathcal{P}^0$ the classical law of the iterated logarithm [11, p. 232] says that

$$\limsup_{r\to\infty} \frac{\Big|\sum_{j=1}^r Y_j(p)\Big|}{\sqrt{2r\log\log r}} = \sqrt{\operatorname{Var} Y_j(p)} = \|p\|_4^4 - 1, \qquad \text{a.s.}$$

Choosing $p(x) = D^{-1/2}\sum_{k\in[-M,M]^d\cap\mathbb{Z}^d} e^{2\pi i k\cdot x}$, we have $\|p\|_2 = 1$ and the elementary estimate $\|p\|_4 \ge \frac{2}{\pi}D^{1/4}$. So

$$\limsup_{r\to\infty} \frac{\sup_{p\in\mathcal{P}^0} |\sum_{j=1}^r [|p(x_j)|^2 - r]|}{\sqrt{r\log\log r}} \ge \Big(\frac{2}{\pi}\Big)^4 D - 1\,.$$

The conclusion follows as in the proof of Theorem 4.3 by applying Kolmogorov's zero-one law. ■

This result can be summarized by saying that for large enough r (r depending on ω) we always have the sampling inequality

$$(48)\quad (r-c\sqrt{r\log\log r})\|p\|_2^2 \le \sum_{j=1}^{r} |p(x_j)|^2 \le (r+c\sqrt{r\log\log r})\|p\|_2^2 \qquad \forall p \in \mathcal{P}_M\,.$$

The condition number of the random matrix $\mathcal{T}$ is therefore

$$\kappa \le (r + c\sqrt{r\log\log r})/(r - c\sqrt{r\log\log r}) \approx 1 + 2c\Big(\frac{\log\log r}{\sqrt{r}}\Big)^{1/2}$$

almost surely for some constant c of order D.

6. A Universal Sampling Theorem and Examples

The main statements (Theorems 4.2, 5.1, Cor. 5.4) reach similar conclusions. At first glance, Theorem 4.2 seems preferable because of its elementary proof and the explicit constants. In this section we focus on the merits of the metric entropy method. This method is extremely flexible and works for many other sampling problems. We formulate a general framework for finite-dimensional sampling theorems and derive a universal sampling theorem in the style of Corollary 5.4. We then will discuss several examples of practical interest.

To begin, we note that the proofs of Theorem 5.1 and Corollary 5.4 do not use any specific properties of trigonometric polynomials. In fact, we have used only the following (interrelated) properties of $\mathcal{P}_M$.

(a) The space $\mathcal{P}_M$ is finite-dimensional and possesses a basis of continuous functions.

(b) All norms on $\mathcal{P}_M$ are equivalent; in the proofs we have used the norms $\|p\|_2, \|p\|_4, \|p\|_\infty$ and $\|\mathbf{a}\|_2$ and the associated equivalence constants. As a consequence the RVs related to the samples $|p(x_j)|^2$ satisfy the uniform estimates of Lemma 5.3.

(c) The unit ball of $\mathcal{P}_M$ is compact. This fact enables the construction of the δ-nets $\mathcal{A}(\delta)$ and suitable estimates for their cardinality.

It is evident that Theorem 5.1 and Corollary 5.4 can be obtained under much more general conditions.

A General Framework. We make the following assumptions.

1. Let $S \subseteq \mathbb{R}^d$ be a compact set and let ν a probability measure on S with $\operatorname{supp}\nu = S$.

2. Let $\mathcal{B}$ be a finite-dimensional subspace of $L^2(S,\nu)$ with a basis $\{e_k : k = 1,\dots,D\}$ of continuous functions. Often this basis is chosen as a finite subset of a Riesz basis for $L^2(S,\nu)$ and in this sense $\mathcal{B}$ may be interpreted as a space of "band-limited" functions in $L^2(S,\nu)$. Since $p = \sum_{k=1}^D a_k e_k$ for every $p \in \mathcal{B}$, all functions in $\mathcal{B}$ are continuous.

The Sampling Problem in $\mathcal{B}$. The task is now to interpolate or to approximate a given data set $\{(x_j, p(x_j)) : j = 1,\dots,r\}$ by a function in $\mathcal{B}$. As in Section 2 this amounts to solving the system of linear equations

$$\sum_{k=1}^{D} a_k e_k(x_j) = p(x_j) = y_j \qquad j = 1, \dots, r\,.$$

Let $\mathcal{U}_{jk} = e_k(x_j)$ and

$$\mathcal{T}_{kl} = (\mathcal{U}^*\mathcal{U})_{kl} = \sum_{j=1}^{r} \overline{e_k(x_j)} e_l(x_j)\,; \tag{49}$$

then we need to solve either the $r \times D$ system

$$\mathcal{U}\mathbf{a} = \mathbf{y}$$

or the $D \times D$ normal equations

$$\mathcal{T}\mathbf{a} = \mathcal{U}^*\mathcal{U}\mathbf{a} = \mathcal{U}^*\mathbf{y}\,.$$

Assume that we can prove the sampling inequality

$$A\|p\|_{2,\nu}^2 \le \sum_{j=1}^{r} |p(x_j)|^2 = \langle \mathcal{T}\mathbf{a}, \mathbf{a}\rangle \le B\|p\|_{2,\nu}^2 \quad \forall p \in \mathcal{B}\,. \tag{50}$$

Inserting the norm equivalence $\alpha\|\mathbf{a}\|_2 \le \|p\|_{2,\nu} \le \beta\|\mathbf{a}\|_2$, (50) then implies the estimates

$$\kappa(\mathcal{T}) \le \frac{\beta^2 B}{\alpha^2 A} \qquad \text{and} \qquad \kappa(\mathcal{U}) \le \Big(\frac{\beta^2 B}{\alpha^2 A}\Big)^{1/2} \tag{51}$$

for the condition numbers of these matrices. Furthermore, $p \in \mathcal{B}$ is uniquely determined by its samples, if and only if $\mathcal{T}$ is invertible, or if and only if $r \ge D$ and $\mathcal{U}$ has full rank.

We can now formulate our main theorem for random sampling in finite-dimensional spaces of band-limited functions.

Theorem 6.1. *If $\{x_j : j \in \mathbb{N}\}$ is a sequence of i.i.d. random variables and if each x_j is ν-distributed over S, then there exist constants $A, B > 0$ depending on S, ν, and D, such that for all $\mu \in (0,1)$, the sampling inequality*

$$(1-\mu) r\|p\|_{2,\nu}^2 \le \sum_{j=1}^{r} |p(x_j)|^2 \le (1+\mu) r\|p\|_{2,\nu}^2 \qquad \forall p \in \mathcal{B} \tag{52}$$

holds with probability at least

$$1 - Ae^{-Br\frac{\mu^2}{1+\mu}}.$$

With the same probability estimate the matrix $\mathcal{T}$ has condition number $\kappa(\mathcal{T}) \le \frac{\beta^2(1+\mu)}{\alpha^2(1-\mu)}$ and also $\kappa(\mathcal{U}) \le \Big(\frac{\beta^2(1+\mu)}{\alpha^2(1-\mu)}\Big)^{1/2}$

Proof. We have already done all the work when we proved Theorem 5.1 and Corollary 5.4. The only minor modifications occur in the constants in Lemma 5.3 and in Step 1 of the proof. We now use the RVs $Y_j(p) = |p(x_j)|^2 - \|p\|_{2,\nu}^2 = |p(x_j)|^2 - \mathbb{E}\,[|p(x_j)|^2]$. ∎

We present the following examples where the general hypotheses are satisfied and so Theorem 6.1 is applicable. Each example yields a new result on random sampling. In some of these examples it seems to be extremely difficult to derive quantitative deterministic results in the style of Theorem 4.1.

Example 1. *Trigonometric Polynomials Revisited.*
Choose a closed set $S \subseteq [0,1]^d$ of positive Lebesgue measure and a probability measure ν with $\operatorname{supp} \nu = S$ and equivalent to λ on S. If $p \in \mathcal{P}_M$ vanishes on S, then by Lemma 3.1 $p \equiv 0$ and consequently $\|p\chi_S\|_{2,\nu} = \left(\int_S |p(x)|^2 \, d\nu(x)\right)^{1/2}$ is equivalent to the L^2-norm on $\mathcal{P}_M$, i.e., there exist constants $\alpha, \beta > 0$ such that

$$\alpha\|p\|_2 \le \|p\chi_S\|_{2,\nu} \le \beta\|p\|_2 \qquad \forall p \in \mathcal{P}_M \, .$$

We state the conclusion of Theorem 6.1 explicitly.

Theorem 6.2. *Suppose that $\{x_j : j \in \mathbb{N}\} \subseteq S$ is a sequence of i.i.d. random variables that are ν-distributed over S. Then there exist constants $A, B > 0$ depending on S, ν and D, such that for all $\mu \in (0,1)$ the sampling inequality*

$$\alpha^2(1-\mu)r\|p\|_2^2 \le \sum_{j=1}^{r} |p(x_j)|^2 \le \beta^2\,(1+\mu)r\|p\|_2^2 \qquad \forall p \in \mathcal{P}_M \tag{53}$$

holds with probability at least

$$1 - Ae^{-Br\frac{\mu^2}{1+\mu}}.$$

With the same probability estimate we have $\kappa(\mathcal{T}) \le \frac{\beta^2(1+\mu)}{\alpha^2(1-\mu)}$.

Comparing with Theorem 5.1 we have been able to change the distribution of the RVs x_j and the target set S in which the samples are taken.

Example 2. *Almost Periodic Functions and Trigonometric Polynomials with Arbitrary Frequencies.* Assume that $S \subseteq \mathbb{R}^d$ is compact and has positive Lebesgue measure and that ν is equivalent to λ on S. Choose exponentials $e^{i\lambda_k \cdot x}$ with arbitrary frequencies $\lambda_k \in \mathbb{R}^d$ ($\lambda_k \in \mathbb{Z}^d$ is the case of trigonometric polynomials) and consider the subspace of almost periodic functions (trigonometric polynomials) on S

$$\mathcal{B} = \{p \in L^2(S) : p(x) = \sum_{k=1}^{D} a_k e^{i\lambda_k \cdot x} \; \chi_S(x) \, .$$

Then Theorem 6.1 applies.

Example 3. *Algebraic Polynomials.* Again assume that $S \subseteq \mathbb{R}^d$ has positive Lebesgue measure and that ν is equivalent to λ on S. Choose a finite set $F \subseteq (\mathbb{N} \cap \{0\})^d$ and consider the space of algebraic polynomials on a compact set $S \subseteq \mathbb{R}^d$ defined as

$$\mathcal{P}_F = \{p \in L^2(S) : p(x) = \sum_{k \in F} a_k x^\alpha \; \chi_S(x)$$

Thus Theorem 6.1 applies also to algebraic polynomials of several variables.

Example 4. *Local Shift-Invariant Spaces.* Let ϕ be a continuous function on $\mathbb{R}^d$ with $\operatorname{supp}\phi \subseteq [-\sigma,\sigma]^d \subseteq S$. The local shift-invariant space $V(\phi, S)$ is defined by

$$V(\phi, S) = \{f \in L^2(S) : f(x) = \sum_{k\in(S+[-\sigma,\sigma]^d)\cap\mathbb{Z}^d} a_k\phi(x-k)\}$$

If we assume that $0 < a \leq \sum_{k\in\mathbb{Z}^d} |\hat{\phi}(\omega-k)|^2 \leq b$ for all $\omega \in \mathbb{R}^d$, then the translates $\phi(x-k), k \in \mathbb{Z}^d$, form a Riesz basis for the generated subspace, and so any finite subset is linearly independent. Thus Theorem 6.1 applies. In dimension $d = 1$ and for certain "generators" ϕ this model is well-understood both numerically [21] and theoretically [1]. In dimension $d > 1$, however, there are no quantitative deterministic estimates. Theorem 6.1 gives the first hint that the numerical methods of [21] also work in higher dimensions. See [2] for a survey of sampling in shift-invariant spaces.

Example 5. *Sampling on the Sphere and Spherical Harmonics.*
Let $S_d = \{x \in \mathbb{R}^{d+1} : |x| = 1\}$ be the unit sphere in $\mathbb{R}^{d+1}$ with surface measure ν_d. We choose the sequence J_ℓ of suitably normalized spherical harmonics [34] as an orthonormal basis for $L^2(S_d, \nu_d)$ and consider the space of band-limited functions on the sphere, namely

$$\mathcal{B} = \{p \in L^2(S_d, \nu_d) : p = \sum_{\ell=1}^{D} a_\ell J_\ell\}\,.$$

Then the conclusions of Theorem 6.1 hold for every sequence of i.i.d. RVs x_j on S_d with x_j being ν_d-distributed.

REMARK: Whereas the asymptotic results for the distribution number hold universally in finite-dimensional vector spaces, the generalization of Theorem 3.2 is more subtle and depends on the support properties of the basis functions. The same proof as in Section 3 shows that the system matrix $\mathcal{T}$ defined in (49) is invertible with probability 1 in Examples 1, 2, and 3 whenever $r \geq D$. On the other hand, for Example 4 it can be shown that $\mathcal{T}$ is always singular with positive probability. As this probability depends on the number of samples r, this observation does not contradict Theorem 6.1.

References

[1] A. Aldroubi and K. Gröchenig. Beurling-Landau-type theorems for non-uniform sampling in shift invariant spline spaces. *J. Fourier Anal. Appl.*, 6(1):93–103, 2000.

[2] A. Aldroubi and K. Gröchenig. Nonuniform sampling and reconstruction in shift-invariant spaces. *SIAM Rev.*, 43(4):585–620, 2001.

[3] A. Averbuch, R. Coifman, D. Donoho, M. Israeli, and S. Y. Fast slant stack: A notion of Radon transform for data in a Cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible. *SIAM J. of Scientific Computing*, To appear.

[4] R. F. Bass. Law of the iterated logarithm for set-indexed partial sum processes with finite variance. *Z. Wahrsch. Verw. Gebiete*, 70(4):591–608, 1985.

[5] R. F. Bass. *Probabilistic techniques in analysis.* Probability and its Applications (New York). Springer-Verlag, New York, 1995.

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

[7] 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.

[8] G. Beylkin. On the fast Fourier transform of functions with singularities. *Appl. Comput. Harmon. Anal.*, 2(4):363–381, 1995.

[9] 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.

[10] 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.

[11] K. L. Chung. *A course in probability theory*. Academic Press, New York-London, second edition, 1974. Probability and Mathematical Statistics, Vol. 21.

[12] R. M. Dudley. Sample functions of the Gaussian process. *Ann. Probability*, 1(1):66–103, 1973.

[13] A. Dutt and V. Rokhlin. Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.*, 14(6):1368–1393, 1993.

[14] A. Dutt and V. Rokhlin. Fast Fourier transforms for nonequispaced data. II. *Appl. Comput. Harmon. Anal.*, 2(1):85–100, 1995.

[15] H. Fassbender. On numerical methods for discrete least-squares approximation by trigonometric polynomials. *Math. Comp.*, 66(218):719–741, 1997.

[16] H. G. Feichtinger, K. Gröchenig, and T. Strohmer. Efficient numerical methods in non-uniform sampling theory. *Numer. Math.*, 69(4):423–440, 1995.

[17] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

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

[19] K. Gröchenig. Irregular sampling, Toeplitz matrices, and the approximation of entire functions of exponential type. *Math. Comp.*, 68(226):749–765, 1999.

[20] K. Gröchenig. Non-uniform sampling in higher dimensions: from trigonometric polynomials to bandlimited functions. In *Modern sampling theory*, Appl. Numer. Harmon. Anal., pages 155–171. Birkhäuser Boston, Boston, MA, 2001.

[21] K. Gröchenig and H. Schwab. Fast local reconstruction methods for nonuniform sampling in shift invariant spaces. *SIAM J. Matrix Anal.*, to appear.

[22] K. Gröchenig and T. Strohmer. Numerical and theoretical aspects of non-uniform sampling of band-limited images. In F. Marvasti, editor, *Nonuniform Sampling: Theory and Applications*, chapter 6, pages 283 – 324. Kluwer, 2001.

[23] S. Kochen and C. Stone. A note on the Borel-Cantelli lemma. *Illinois J. Math.*, 8:248–251, 1964.

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

[25] Y. I. Lyubarskii and K. Seip. Complete interpolating sequences for Paley-Wiener spaces and Muckenhoupt's (A_p) condition. *Rev. Mat. Iberoamericana*, 13(2):361–376, 1997.

[26] B. S. Pavlov. The basis property of a system of exponentials and the condition of Muckenhoupt. *Dokl. Akad. Nauk SSSR*, 247(1):37–40, 1979.

[27] D. Potts and G. Steidl. Fourier reconstruction of functions from their nonstandard sampled Radon transform. *J. Fourier Anal. Appl.*, 8(6):513–533, 2002.

[28] 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.

[29] M. Rauth and T. Strohmer. Smooth approximation of potential fields from noisy scattered data. *Geophysics*, 63(1):85–94, 1998.

[30] L. Reichel, G. S. Ammar, and W. B. Gragg. Discrete least squares approximation by trigonometric polynomials. *Math. Comp.*, 57(195):273–289, 1991.

[31] K. Seip and A. M. Ulanovskii. Random exponential frames. *J. London Math. Soc. (2)*, 53(3):560–568, 1996.

[32] Y. Shkolnisky, A. Averbuch, M. Israeli, R. Coifman, and D. Donoho. 2d Fourier based discrete Radon transform. *Preprint.*

[33] G. Steidl. A note on fast Fourier transforms for nonequispaced grids. *Adv. Comput. Math.*, 9(3-4):337–352, 1998.

[34] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces.* Princeton Univ. Press, Princeton, NJ, 1971. Princeton Mathematical Series, No. 32.

[35] T. Strohmer. Computationally attractive reconstruction of band-limited images from irregular samples. *IEEE Trans. on Image Processing*, 6(4):540–548, 1997.

[36] T. Strohmer. Numerical analysis of the non-uniform sampling problem. *J. Comput. Appl. Math.*, 122(1-2):297–316, 2000. Numerical analysis 2000, Vol. II: Interpolation and extrapolation.

[37] T. Strohmer, T. Binder, and M. Süssner. How to recover smooth object boundaries in noisy medical images. In *Proc. ICIP'96*, pages 331–334, 1996.

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

E-mail address: BASS@MATH.UCONN.EDU,GROCH@MATH.UCONN.EDU