Probabilistic Spherical Marcinkiewicz-Zygmund Inequalities

Albrecht Böttcher Stefan Kunis Daniel Potts

Recently, norm equivalences between spherical polynomials and their sample values at scattered sites have been proved. These so-called Marcinkiewicz-Zygmund inequalities involve a parameter that characterizes the density of the sampling set and they are applicable to all polynomials whose degree does not exceed an upper bound that is determined by the density parameter. We show that if one is satisfied by norm equivalences that hold with prescribed probability only, then the upper bound for the degree of the admissible polynomials can be enlarged significantly and that then, moreover, there exist fixed sampling sets which work for polynomials of all degrees.

1 Introduction

Let $f(x) = \sum_{k=-N}^{N} \hat{f}_k \mathrm{e}^{\mathrm{i}kx}$ be a trigonometric polynomial of degree N, denote by $\boldsymbol{f}$ the vector

$$\boldsymbol{f} = \left(f\left(\frac{2\pi j}{2N+1} \right) \right)_{j=0}^{2N},$$

define the weight $\boldsymbol{W} = (w_j)_{j=1}^{2N}$ by $w_j = 2\pi/(2N+1)$, and put

$$\|\boldsymbol{f}\|_{\boldsymbol{W},2}^2 := \sum_{j=0}^{2N} \frac{2\pi}{2N+1} \left| f\left(\frac{2\pi j}{2N+1} \right) \right|^2, \quad \|f\|_2^2 := \int_0^{2\pi} |f(\xi)|^2 \mathrm{d}\xi.$$

Parseval's equality implies that $\|f\|_2^2 = 2\pi \sum_{k=-N}^{N} |\hat{f}_k|^2$, and since the Fourier matrix $\boldsymbol{U} = (2N+1)^{-1/2} \left(\mathrm{e}^{2\pi \mathrm{i} jk/(2N+1)} \right)_{j,k=0}^{2N}$ is unitary, it follows that $\|\boldsymbol{f}\|_{\boldsymbol{W},2}^2 = 2\pi \sum_{k=-N}^{N} |\hat{f}_k|^2$.

AMS subject classification. 41A17, 33C55, 60H30, 15A60.
Keywords and Phrases. Scattered data, Marcinkiewicz-Zygmund inequalities, Spherical harmonics, Random polynomials.
Chemnitz Universitiy of Technology, Department of Mathematics, 09107 Chemnitz, Germany, {aboettch,kunis,potts}@mathematik.tu-chemnitz.de

We therefore arrive at the equality $\|\boldsymbol{f}\|_{\boldsymbol{W},2} = \|f\|_2$. This well known observation is the origin of so-called Marcinkiewicz-Zygmund inequalities, which state that, under certain assumptions,

$$(1-\varepsilon)\|f\|_p \leq \|\boldsymbol{f}\|_{\boldsymbol{W},p} \leq (1+\varepsilon)\|f\|_p. \tag{1}$$

Here $\|\cdot\|_p$ is the L^p norm, $\|\cdot\|_{\boldsymbol{W},p}$ denotes a weighted ℓ^p norm, f is given on some manifold, and $\boldsymbol{f}$ is the vector of samples of f at certain points, the sampling nodes, on the manifold. We remark that inequality (1) gives bounds on the sampling operator and its inverse, which is of importance in connection with the stable reconstruction of polynomials from their samples and the design of quadrature rules.

In the case of trigonometric polynomials, Marcinkiewicz-Zygmund inequalities for equispaced samples were established in [17, 16]. However, measurements are typically taken nonuniformly and, moreover, there do not exist equidistributed sampling sets of sufficiently high cardinality on the unit spheres $\mathbb{S}^d$ for $d \geq 2$. This motivates the increasing interest in norm equivalences for less regular sampling sets. At least since [7], one knows sharp versions of L^2 norm equivalences for trigonometric polynomials under the assumption that the sampling set contains no holes larger than the inverse polynomial degree. Large sieve estimates give upper bounds for nonequispaced sampled trigonometric polynomials, see e.g. [13, 6], and results for randomly chosen sampling nodes were obtained in [1, 18, 8, 5].

The passage from trigonometric polynomials living on the unit circle $\mathbb{S}^1$ or the torus $\mathbb{S}^1 \times \ldots \times \mathbb{S}^1$ to spherical harmonics on the unit sphere $\mathbb{S}^d$ is not trivial. In recent years, L^∞ and general L^p Marcinkiewicz-Zygmund inequalities on spheres were proved in [9] and [14, 4, 5], respectively. The $L^2(\mathbb{S}^2)$ case was studied in more detail and tested numerically in [10].

The Marcinkiewicz-Zygmund inequalities cited above guarantee (1) for all spherical harmonics of degree at most N on the unit sphere $\mathbb{S}^d$ provided N satisfies $N \leq \varepsilon/(RB_d)$ where R is a partition norm associated with the sampling set and B_d is a constant depending only on d. Known estimates deliver quite large values for B_d, which results in a severe limitation for N. The purpose of this paper is to reveal that for given R one can extend the admissible polynomial degrees spectacularly if one admits to have (1) for $p = 2$ and with a certain prescribed probability only. To be more precise, we assume that the coefficients in the linear combinations of the spherical harmonics of degree at most N on $\mathbb{S}^d$ are taken at random from the uniform distribution on a ball of appropriate dimension. Under this assumption we show in Section 3 that if $\varepsilon \in (0,1)$ and $\eta \in (0,1)$ are given, then we can find a $\varrho_0 > 0$ such that if $R < \varrho_0$ then

$$\mathbb{P}\big[(1-\varepsilon)\|f\|_2 \leq \|\boldsymbol{f}\|_{\boldsymbol{W},2} \leq (1+\varepsilon)\|f\|_2\big] \geq 1-\eta \tag{2}$$

whenever $N \leq 1/R$. Moreover, in Section 4 we prove that there exist fixed sampling sets and weights such that (2) is true *for every* $N \geq 0$.

2 Preliminaries

2.1 Sampling data

Let $d \geq 1$ and $\mathbb{S}^d := \{\boldsymbol{\xi} \in \mathbb{R}^{d+1} : |\boldsymbol{\xi}|^2 = 1\}$, where $|\cdot|$ is the usual Euclidean norm. Throughout what follows we assume that we are given a finite set $\mathcal{R} := \{R_1, \ldots, R_M\}$ of closed and nonoverlapping regions R_j of $\mathbb{S}^d$ such that $\cup_{j=1}^M R_j = \mathbb{S}^d$ and a finite set $\mathcal{X} := \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M\}$ of points $\boldsymbol{\xi}_j$ such that $\boldsymbol{\xi}_j$ is in the interior of R_j. Of course, nonoverlapping means that R_i and R_j have no common inner points for $i \neq j$. We refer to the points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M$ as the sampling nodes and to $(\mathcal{R}, \mathcal{X})$ as a sampling pair. Note that we require that each R_j contains exactly one sampling node. We denote by μ_d the usual measure on $\mathbb{S}^d$, define $w_j := \mu_d(R_j)$, and call $\boldsymbol{W} := \{w_1, \ldots, w_M\}$ the weight. Clearly, $\sum_j w_j = \mu_d(\mathbb{S}^d)$. The partition norm $R = \|\mathcal{R}\|$ is defined as

$$R := \max_j \operatorname{diam} R_j := \max_j \max_{\boldsymbol{\xi}, \boldsymbol{\eta} \in R_j} d(\boldsymbol{\xi}, \boldsymbol{\eta}),$$

where $d(\boldsymbol{\xi}, \boldsymbol{\eta}) := \arccos(\boldsymbol{\xi} \cdot \boldsymbol{\eta})$ is the geodesic distance between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. The separation distance q of the set $\mathcal{X}$ is

$$q := \min_{j \neq \ell} d(\boldsymbol{\xi}_j, \boldsymbol{\xi}_\ell).$$

Note that the partition norm, the separation distance, and the frequently used so-called mesh norm $\|\mathcal{X}\| := \max_{\boldsymbol{\xi}} \min_j d(\boldsymbol{\xi}, \boldsymbol{\xi}_j)$ are related by $q \leq \|\mathcal{X}\| \leq R \leq 2\|\mathcal{X}\|$. Obviously, there is a constant C_2 depending only on d such that

$$w_j \leq C_2 R^d, \tag{3}$$

which after summing up implies that $\mu_d(\mathbb{S}^d) \leq C_2 M R^d$. The M spherical disks $D_j := \{\boldsymbol{\xi} \in \mathbb{S}^d : d(\boldsymbol{\xi}, \boldsymbol{\xi}_j) \leq q/2\}$ do not overlap and there is a constant C_1 depending only on d such that $\mu_d(D_j) \geq C_1 q^d$. The last inequality and (3) imply that

$$C_1 M q^d \leq \mu_d(\mathbb{S}^d) \leq C_2 M R^d. \tag{4}$$

Finally notice that $\mu_d(\mathbb{S}^d) = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$.

If only the sampling nodes $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M$ are given, there are a variety of ways to construct a partition $\mathcal{R}$ as above such that each $\boldsymbol{\xi}_j$ belongs to exactly one R_j. Often one takes the so-called Voronoi partition, which is determined by

$$R_j = \{\boldsymbol{\xi} \in \mathbb{S}^d : d(\boldsymbol{\xi}, \boldsymbol{\xi}_j) \leq \min_\ell d(\boldsymbol{\xi}, \boldsymbol{\xi}_\ell)\}.$$

For $f \in C(\mathbb{S}^d)$, we denote by $\boldsymbol{f}$ the vector $(f(\boldsymbol{\xi}_j))_{j=1}^M$. The L^p norm of f is given by

$$\|f\|_p^p := \int_{\mathbb{S}^d} |f(\boldsymbol{\xi})|^p \mathrm{d}\mu_d(\boldsymbol{\xi}) \quad (1 \leq p < \infty), \quad \|f\|_\infty := \max_{\boldsymbol{\xi} \in \mathbb{S}^d} |f(\boldsymbol{\xi})|$$

and the $\ell^p_{\boldsymbol{W}}$ norm of a vector $\boldsymbol{g} = (g_j)_{j=1}^M \in \mathbb{C}^M$ is defined by

$$\|\boldsymbol{g}\|_{\boldsymbol{W},p}^p := \sum_{j=1}^M w_j |g_j|^p \quad (1 \leq p < \infty), \quad \|\boldsymbol{g}\|_{\boldsymbol{W},\infty} := \max_j |g_j|.$$

2.2 Spherical harmonics

We refer to [15, 20] for a thorough introduction to spherical harmonics and here confine ourselves to the following. We denote the spherical harmonics of degree at most N on $\mathbb{S}^d$ by $\{Y_k^d\}_{k=1}^{\mathcal{N}_d(N)}$. These span a subspace Π_N^d of $L^2(\mathbb{S}^d, \mathrm{d}\mu_d)$ and are an orthonormal basis in this subspace. The dimension of Π_N^d is

$$\mathcal{N}_d(N) = \frac{(2N+d)\Gamma(N+d)}{\Gamma(d+1)\Gamma(N+1)} \sim \frac{2}{\Gamma(d+1)}\, N^d, \tag{5}$$

where $x_N \sim y_N$ means that $x_N/y_N \to 1$ as $N \to \infty$. The $\mathcal{N}_1(N) = 2N+1$ spherical harmonics of degree at most N are just the trigonometric polynomials $\{(1/\sqrt{2\pi})\mathrm{e}^{\mathrm{i}kx}\}_{k=-N}^N$. Notice that

$$\mathcal{N}_2(N) = (N+1)^2, \quad \mathcal{N}_3(N) = \frac{1}{6}(N+1)(N+2)(2N+3).$$

A finer decomposition of Π_N^d is as follows. Let H_κ^d be the spherical harmonics whose degree is exactly κ. We label the spherical harmonics in H_κ^d by $\{Y_{\kappa,i}^d\}_{i=1}^{\mathcal{H}_d(\kappa)}$. The dimension of H_κ^d is known to be

$$\mathcal{H}_d(\kappa) = \frac{(2\kappa+d-1)\Gamma(\kappa+d-1)}{\Gamma(d)\Gamma(\kappa+1)} \le H_d \kappa^{d-1} \tag{6}$$

with some constant $H_d \in (0,\infty)$ depending only on d. Finally, let $\mathsf{C}_\kappa^{(d-1)/2} : [-1,1] \to \mathbb{R}$ be the $(\kappa, \frac{d-1}{2})$th Gegenbauer polynomial determined by the normalization $\mathsf{C}_\kappa^{(d-1)/2}(1) = 1$. Herglotz' famous addition theorem says that

$$\sum_{i=1}^{\mathcal{H}_d(\kappa)} Y_{\kappa,i}^d(\boldsymbol{\xi})\overline{Y_{\kappa,i}^d(\boldsymbol{\eta})} = \frac{\mathcal{H}_d(\kappa)}{\mu_d(\mathbb{S}^d)} \mathsf{C}_\kappa^{(d-1)/2}(\boldsymbol{\xi}\cdot\boldsymbol{\eta}) \tag{7}$$

for all $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{S}^d$. Taking into account that $\sum_{\kappa=0}^N \mathcal{H}_d(\kappa) = \mathcal{N}_d(N)$, we obtain in particular that

$$\sum_{k=1}^{\mathcal{N}_d(N)} |Y_k^d(\boldsymbol{\xi})|^2 = \sum_{\kappa=0}^{N}\sum_{i=1}^{\mathcal{H}_d(\kappa)} |Y_{\kappa,i}^d(\boldsymbol{\xi})|^2 = \frac{\mathcal{N}_d(N)}{\mu_d(\mathbb{S}^d)} \tag{8}$$

for all $\boldsymbol{\xi} \in \mathbb{S}^d$. We will also make use of the following estimate, which can be found in [19, Theorem 2.9]: if $0 < \varphi < \pi$, then

$$\left|\mathsf{C}_\kappa^{(d-1)/2}(\cos\varphi)\right| \le G_d \kappa^{-(d-1)/2} |\sin\varphi|^{-(d-1)/2}, \tag{9}$$

where $G_d \in (0,\infty)$ depends only on d.

2.3 Two deterministic inequalities

The following results are from [4]. If $NR \le 1$, then

$$\|\boldsymbol{f}\|_{\boldsymbol{W},1} \le (1 + B_d NR)\|f\|_1 \tag{10}$$

for all $f \in \Pi_N^d$ where B_d is a constant that depends only on d and is bounded by $B_d \le (2\sqrt{3})^d(5d+1)$. Moreover, if $NR \le \varepsilon/B_d$ with $\varepsilon \in (0,1)$, then

$$(1-\varepsilon)\|f\|_p \le \|\boldsymbol{f}\|_{\boldsymbol{W},p} \le (1+\varepsilon)\|f\|_p \tag{11}$$

for all $1 \le p \le \infty$ and all $f \in \Pi_N^d$.

2.4 Probability

We denote by $\mathbb{P}(X)$ the probability of an event X. Let $\mathbb{C}^k$ be equipped with the ℓ^2 norm, $\|\boldsymbol{x}\|_2^2 = \sum_{j=1}^k |x_j|^2$, and let $\mathbb{B}^k := \{\boldsymbol{x} \in \mathbb{C}^k : \|\boldsymbol{x}\|_2^2 = 1\}$. The spectral and Frobenius norm of a matrix $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ will be denoted by $\|\boldsymbol{A}\|_2$ and $\|\boldsymbol{A}\|_{\mathrm{F}}$, respectively. We start with the following result.

Lemma 2.1. *If $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{x}$ is drawn randomly from the uniform distribution on $\mathbb{B}^n$, then the expectation and variance of the random variable $\|\boldsymbol{A}\boldsymbol{x}\|_2^2/\|\boldsymbol{x}\|_2^2$ are*

$$E\left(\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2}\right) = \frac{\|\boldsymbol{A}\|_{\mathrm{F}}^2}{n}, \quad \sigma^2\left(\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2}\right) = \frac{2}{n+2}\left(\frac{\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n} - \left(\frac{\|\boldsymbol{A}\|_{\mathrm{F}}^2}{n}\right)^2\right). \tag{12}$$

Proof. This was essentially established in [2]. There square matrices $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ were considered and it was shown that the expectation and variance are

$$\frac{s_1^2 + \ldots + s_n^2}{n} \quad \text{and} \quad \frac{2}{n+2}\left(\frac{s_1^4 + \ldots + s_n^4}{n} - \left(\frac{s_1^2 + \ldots + s_n^2}{n}\right)^2\right), \tag{13}$$

respectively, where $s_1, \ldots, s_n$ are the singular values of $\boldsymbol{A}$. For rectangular matrices, $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, (13) becomes

$$\frac{s_1^2 + \ldots + s_k^2}{n} \quad \text{and} \quad \frac{2}{n+2}\left(\frac{s_1^4 + \ldots + s_k^4}{n} - \left(\frac{s_1^2 + \ldots + s_k^2}{n}\right)^2\right), \tag{14}$$

and was already used in [3]. Here $k = \min(m,n)$ and $s_1, \ldots, s_k$ are the singular values of $\boldsymbol{A}$. Note that if $m < n$, then (14) follows from the square case by considering the square matrix $(\boldsymbol{A}^\top \boldsymbol{0})^\top \in \mathbb{C}^{n\times n}$, while if $m > n$, then (14) can be derived from the square case by writing $\boldsymbol{A} = \boldsymbol{U}(\boldsymbol{B}^\top \boldsymbol{0})^\top \boldsymbol{V}^\top$ with $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ and unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in $\mathbb{C}^{m\times m}$. If we denote by $\lambda_1, \ldots, \lambda_m$ the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^*$, then

$$s_1^2 + \ldots + s_k^2 = \|\boldsymbol{A}\|_{\mathrm{F}}^2, \quad s_1^4 + \ldots + s_k^4 = \lambda_1^2 + \ldots + \lambda_m^2 = \|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2,$$

which gives (12). $\square$

In what follows we will employ the following consequence of Lemma 2.1.

Corollary 2.2. *Let $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and suppose $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = n$. If $\boldsymbol{x}$ is taken at random from the uniform distribution on $\mathbb{B}^n$, then*

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \le 1+\varepsilon\right) \ge 1 - \frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n^2\varepsilon^2(2-\varepsilon)^2} \tag{15}$$

for every $\varepsilon \in (0,1)$.

Proof. Put $\delta = \varepsilon(2-\varepsilon)$. Lemma 2.1 and Chebyshev's inequality imply that

$$\begin{aligned}\mathbb{P}\left(1-\delta \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \le 1+\delta\right) &\ge 1 - \frac{1}{\delta^2}\sigma^2\left(\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2}\right)\\ \ge 1 - \frac{1}{\delta^2}\,\frac{2}{n}\left(\frac{\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n} - 1\right) &\ge 1 - \frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n^2\delta^2}.\end{aligned} \tag{16}$$

Since $1-\delta = (1-\varepsilon)^2$ and $1+\delta \le (1+\varepsilon)^2$, it follows that

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \le 1+\varepsilon\right) \ge \mathbb{P}\left(1-\delta \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \le 1+\delta\right),$$

which together with (16) yields (15). □

To convey to the reader a feeling of what the use of (15) is, let $\boldsymbol{A}$ be an $n\times n$ matrix with $\|\boldsymbol{A}\|_2 \le \gamma$ and $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = n$. Since

$$\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2 \le \|\boldsymbol{A}\|_{\mathrm{F}}^2\|\boldsymbol{A}^*\|_2^2 = n\,\|\boldsymbol{A}\|_2^2 \le n\gamma^2, \tag{17}$$

estimate (15) gives

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \le 1+\varepsilon\right) \ge 1 - \frac{2\gamma^2}{n\varepsilon^2} \tag{18}$$

for each $\varepsilon \in (0,1)$. To make things a little more tricky, one can replace the $(2-\varepsilon)^2$ in (15) by 1 to get

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \le 1+\varepsilon\right) \ge 1 - \frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n^2\varepsilon^2} \tag{19}$$

and then choose $\varepsilon = 1/\sqrt[3]{n}$ to conclude that

$$\mathbb{P}\left(1-\frac{1}{\sqrt[3]{n}} \le \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \le 1+\frac{1}{\sqrt[3]{n}}\right) \ge 1 - \frac{2\gamma^2}{\sqrt[3]{n}}. \tag{20}$$

Consequently, although deterministically we cannot say more than $\|\boldsymbol{A}\boldsymbol{x}\|_2 \le \gamma\|\boldsymbol{x}\|_2$, in fact the values of $\|\boldsymbol{A}\boldsymbol{x}\|_2$ are concentrated tightly close to $\|\boldsymbol{x}\|_2$ with a probability that converges to 1 as the matrix dimension goes to infinity. If $\boldsymbol{A}$ is the Fourier matrix, which is unitary, then (18) and (20) are trivial because $\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ for all $\boldsymbol{x}$. We will see that the matrices emerging in nonuniform sampling problems still enjoy the properties that $\|\boldsymbol{A}\|_2$ has controllable upper bounds and that $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = n$. Thus, for these matrices we have estimates like (18) or (20), which, of course, are useful in connection with Marcinkiewicz-Zygmund inequalities.

3 Sampling pairs for polynomials of high degree

In what follows we always assume that the coefficient vector $(\hat{f}_k)_{k=1}^{\mathcal{N}_d(N)}$ of the spherical polynomial

$$f = \sum_{k=1}^{\mathcal{N}_d(N)} \hat{f}_k Y_k^d \in \Pi_N^d$$

is taken at random from the uniform distribution on the ball $\mathbb{B}^{\mathcal{N}_d(N)}$. The constants B_d are those of Subsection 2.3.

Theorem 3.1. *If $NR \le 1$ then*

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{W},2}}{\|f\|_2} \le 1+\varepsilon\right) \ge 1 - \frac{2(1+B_d NR)}{\mathcal{N}_d(N)\varepsilon^2(2-\varepsilon)^2}$$

for each $\varepsilon \in (0,1)$.

Proof. Equip Π_N^d and $\mathbb{C}^M$ with the L^p and $\ell^p_{\boldsymbol{W}}$ norms, respectively, consider the operator

$$\boldsymbol{B} : \Pi_N^d \to \mathbb{C}^M, \quad f \mapsto \boldsymbol{f} := (f(\boldsymbol{\xi}_1), \dots, f(\boldsymbol{\xi}_M)),$$

and denote by $\|\boldsymbol{B}\|_p$ its norm. Since obviously $\|\boldsymbol{f}\|_{\boldsymbol{W},\infty} \le \|f\|_\infty$, we see that $\|\boldsymbol{B}\|_\infty \le 1$. From (10) we infer that $\|\boldsymbol{B}\|_1 \le 1 + B_d NR$. The Riesz-Thorin interpolation theorem therefore implies that

$$\|\boldsymbol{B}\|_2 \le \|\boldsymbol{B}\|_1^{1/2} \|\boldsymbol{B}\|_\infty^{1/2} \le \sqrt{1+B_d NR}.$$

Consequently,

$$\|\boldsymbol{f}\|_{\boldsymbol{W},2} \le \sqrt{1+B_d NR}\ \|f\|_2 \tag{21}$$

for all $f \in \Pi_N^d$. Now provide Π_N^d and $\mathbb{C}^M$ with the usual (unweighted) L^2 and ℓ^2 norms, respectively, and consider the operator

$$\boldsymbol{A} : \Pi_N^d \to \mathbb{C}^M, \quad f \mapsto (\sqrt{w_1} f(\boldsymbol{\xi}_1), \dots, \sqrt{w_M} f(\boldsymbol{\xi}_M)).$$

From (21) we see that $\|\boldsymbol{A}\|_2 \le \sqrt{1+B_d NR}$ and the addition theorem (8) yields

$$\begin{aligned} \|\boldsymbol{A}\|_{\mathrm{F}}^2 &= \sum_{k=1}^{\mathcal{N}_d(N)} \|\boldsymbol{A} Y_k^d\|_2^2 = \sum_{k=1}^{\mathcal{N}_d(N)} \sum_{j=1}^{M} w_j |Y_k^d(\boldsymbol{\xi}_j)|^2 \\ &= \sum_{j=1}^{M} w_j \sum_{k=1}^{\mathcal{N}_d(N)} |Y_k^d(\boldsymbol{\xi}_j)|^2 = \sum_{j=1}^{M} w_j \frac{\mathcal{N}_d(N)}{\mu_d(\mathbb{S}^d)} = \mathcal{N}_d(N). \end{aligned} \tag{22}$$

Thus, using Corollary 2.2 and taking into account that $\|\boldsymbol{f}\|_{\boldsymbol{W},2} = \|\boldsymbol{A}f\|_2$, we obtain that

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{W},2}}{\|f\|_2} \le 1+\varepsilon\right) \ge 1 - \frac{2\,\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{\mathcal{N}_d(N)^2\varepsilon^2(2-\varepsilon)^2}. \tag{23}$$

Estimating as in (17) we get $\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2 \le \mathcal{N}_d(N)(1+B_d NR)$. Inserting this into the right-hand side of (23), we arrive at the asserted inequality. □

The following corollary tells us in two different ways that the ratio $\|\boldsymbol{f}\|_{\boldsymbol{W},2}/\|f\|_2$ is tightly concentrated around 1 with a probability that converges to 1 as N goes to infinity.

Corollary 3.2. *Let $NR \le 1$. If $0 < \alpha < d/2$, then*

$$a_N := \frac{2(1 + B_d NR)N^{2\alpha}}{\mathcal{N}_d(N)} \sim \frac{\Gamma(d+1)(1 + B_d NR)}{N^{d-2\alpha}}$$

and

$$\mathbb{P}\left(1 - \frac{1}{N^\alpha} \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{w},2}}{\|f\|_2} \le 1 + \frac{1}{N^\alpha}\right) \ge 1 - a_N.$$

If $0 < \beta < d$, then

$$b_N := \sqrt{\frac{2N^\beta(1 + B_d NR)}{\mathcal{N}_d(N)}} \sim \frac{\sqrt{\Gamma(d+1)(1 + b_d NR)}}{N^{(d-\beta)/2}}$$

and

$$\mathbb{P}\left(1 - b_N \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{w},2}}{\|f\|_2} \le 1 + b_N\right) \ge 1 - \frac{1}{N^\beta}.$$

Proof. Use Theorem 3.1 with $(2-\varepsilon)^2$ replaced by 1 (such as in (19)) with $\varepsilon = 1/N^\alpha$ and $\varepsilon = b_N$, respectively, and take into account the asymptotic formula (5). □

Here is another striking consequence of Theorem 3.1.

Corollary 3.3. *Let $\varepsilon \in (0,1)$ and $\eta \in (0,1)$. If*

$$R \le \frac{\varepsilon}{B_d}\left(\frac{\Gamma(d+1)(1+\varepsilon)}{\eta\varepsilon^2(2-\varepsilon)^2}\right)^{-1/d} \tag{24}$$

then

$$1 - \varepsilon \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{w},2}}{\|f\|_2} \le 1 + \varepsilon \tag{25}$$

for $N \le \varepsilon/(B_d R)$ and

$$\mathbb{P}\left(1 - \varepsilon \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{w},2}}{\|f\|_2} \le 1 + \varepsilon\right) \ge 1 - \eta \tag{26}$$

for every $N \le 1/R$.

Proof. From (11) we deduce that the deterministic inequality (25) is true whenever $B_d NR \le \varepsilon$. We are therefore left with proving (26) for $\varepsilon/(B_d R) < N \le 1/R$. Theorem 3.1 tells us that (26) is certainly satisfied if $NR \le 1$ and

$$\frac{2(1 + B_d NR)}{\mathcal{N}_d(N)\varepsilon^2(2-\varepsilon)^2} \le \eta. \tag{27}$$

By formula (5),

$$\mathcal{N}_d(N) = \frac{2N + d}{\Gamma(d+1)}(N + d - 1)\dots(N+1) \ge \frac{2N^d}{\Gamma(d+1)}$$

and hence (27) holds if

$$\frac{\Gamma(d+1)(1+B_dNR)}{N^d\varepsilon^2(2-\varepsilon)^2} \le \eta,$$

or equivalently,

$$N^d - a(1+B_dNR) \ge 0 \tag{28}$$

with $a := \Gamma(d+1)/(\eta\varepsilon^2(2-\varepsilon)^2)$. Consider the function $F(x) = x^d - a(1+B_dRx)$. Our assumption (24) implies that $F(\varepsilon/(B_dR)) \ge 0$. The function $F(x)$ is monotonously increasing if $dx^{d-1} \ge aB_dR$. Consequently, we get (28) for all $N \ge \varepsilon/(B_dR)$ provided we can show that $d(\varepsilon/(B_dR))^{d-1} \ge aB_dR$, which is in turn equivalent to the inequality $aB_d^dR^d \le d\varepsilon^{d-1}$. But (24) gives $aB_d^dR^d \le \varepsilon^d/(1+\varepsilon)$, and since $1/(1+\varepsilon) \le d/\varepsilon$, we arrive at the desired inequality $aB_d^dR^d \le d\varepsilon^{d-1}$. Thus, (28) holds for all $N \ge \varepsilon/(B_dR)$. It follows that in order to guarantee (26) we need nothing but the remaining inequality $NR \le 1$. $\square$.

4 Universal sampling pairs

We abbreviate $\mathcal{N}_d(N)$ to n. Given a sampling pair $(\mathcal{R}, \mathcal{X}) = (R_1, \ldots, R_M; \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M)$, we denote by $\boldsymbol{A}$ the $M \times n$ matrix

$$\boldsymbol{A} = \begin{pmatrix} \sqrt{w_1}\, Y_1^d(\boldsymbol{\xi}_1) & \ldots & \sqrt{w_1}\, Y_n^d(\boldsymbol{\xi}_1) \\ \vdots & & \vdots \\ \sqrt{w_M}\, Y_1^d(\boldsymbol{\xi}_M) & \ldots & \sqrt{w_M}\, Y_n^d(\boldsymbol{\xi}_M) \end{pmatrix}.$$

Clearly, if $f(\boldsymbol{\xi}) = \sum_{k=1}^n \hat{f}_k Y_k^d(\boldsymbol{\xi})$ and $\boldsymbol{x}$ denotes the column $(\hat{f}_k)_{k=1}^n$, then

$$\begin{aligned} &\|\boldsymbol{x}\|_2^2 = \|f\|_2^2, \\ &\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \sum_{j=1}^M w_j \left| \sum_{k=1}^n \hat{f}_k Y_k^d(\boldsymbol{\xi}_j) \right|^2 = \sum_{j=1}^M w_j |f(\boldsymbol{\xi}_j)|^2 = \|f\|_{\boldsymbol{W},2}^2. \end{aligned}$$

We know from (22) that $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = n$. Thus, Corollary 2.2 gives

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{W},2}}{\|f\|_2} \le 1+\varepsilon\right) \ge 1 - \frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n^2\varepsilon^2(2-\varepsilon)^2}. \tag{29}$$

We begin with $\mathbb{S}^1$.

Theorem 4.1. *Let $d = 1$. Given $\varepsilon \in (0,1)$, $\eta \in (0,1)$, and $L \in (1,\infty)$, there exists a positive number $\varrho_0 = \varrho_0(\varepsilon, \eta, L)$ such that*

$$\mathbb{P}\left(1-\varepsilon \le \frac{\|\boldsymbol{f}\|_{\boldsymbol{W},2}}{\|f\|_2} \le 1+\varepsilon\right) \ge 1-\eta \tag{30}$$

for every polynomial degree $N \ge 0$ if only the uniformity condition $R/q < L$ and the density condition $R < \varrho_0$ are satisfied.

Proof. If $d = 1$, then $n = 2N + 1$, $\mathbb{S}^1$ may be identified with the complex unit circle, and $\boldsymbol{A}$ may accordingly be replaced by

$$\boldsymbol{A} = \begin{pmatrix} \sqrt{w_1} & \sqrt{w_1}\,\boldsymbol{\xi}_1 & \dots & \sqrt{w_1}\,\boldsymbol{\xi}_1^{2N} \\ \vdots & \vdots & & \vdots \\ \sqrt{w_M} & \sqrt{w_M}\,\boldsymbol{\xi}_M & \dots & \sqrt{w_M}\boldsymbol{\xi}_M^{2N} \end{pmatrix}.$$

We have $|(AA^*)_{jj}| = w_j(2N + 1)$ and hence (3) shows that

$$\sum_{j=1}^{M} |(AA^*)_{jj}|^2 = (2N + 1)^2 \sum_{j=1}^{M} w_j^2 \le C_2^2 M R^2 (2N + 1)^2. \tag{31}$$

For $j \ne \ell$, we may again use (3) to get

$$|(AA^*)_{j\ell}| = \sqrt{w_j w_\ell} \left| \sum_{k=0}^{2N} \boldsymbol{\xi}_j^k \overline{\boldsymbol{\xi}}_\ell^k \right| = \sqrt{w_j w_\ell} \left| \frac{1 - (\boldsymbol{\xi}_j \overline{\boldsymbol{\xi}}_\ell)^{2N+1}}{1 - \boldsymbol{\xi}_j \overline{\boldsymbol{\xi}}_\ell} \right| \le 2C_2 R \frac{1}{|\boldsymbol{\xi}_j - \boldsymbol{\xi}_\ell|}.$$

Thus, if j is fixed, then

$$\begin{aligned} \sum_{\ell \ne j} |(AA^*)_{j\ell}|^2 &\le 4C_2^2 R^2 \sum_{\ell \ne j} \frac{1}{|\boldsymbol{\xi}_j - \boldsymbol{\xi}_\ell|^2} \\ &\le 4C_2^2 R^2 \cdot 2 \left(\frac{1}{q^2} + \frac{1}{(2q)^2} + \frac{1}{(3q)^2} + \dots \right) \le C_3 \frac{R^2}{q^2}. \end{aligned}$$

Consequently,

$$\sum_{j=1}^{M} \sum_{\ell \ne j} |(AA^*)_{j\ell}|^2 \le C_3 \frac{M R^2}{q^2}. \tag{32}$$

From (31) and (32) we obtain that

$$\frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{(2N + 1)^2 \varepsilon^2 (2 - \varepsilon)^2} \le \frac{2C_2^2 M R^2}{\varepsilon^2 (2 - \varepsilon)^2} + \frac{2C_3 M R^2}{4N^2 q^2 \varepsilon^2 (2 - \varepsilon)^2}. \tag{33}$$

If $RN \le \varepsilon/B_1$, then $1 - \varepsilon \le \|\boldsymbol{f}\|_{\boldsymbol{W},2}/\|f\|_2 \le 1 + \varepsilon$ deterministically. So let $RN > \varepsilon/B_1$. In that case the right-hand side of (33) is at most

$$\frac{2C_2^2 M R^2}{\varepsilon^2 (2 - \varepsilon)^2} + \frac{2C_3 B_1^2 M R^4}{4q^2 \varepsilon^4 (2 - \varepsilon)^2}. \tag{34}$$

The first term in (34) equals

$$\frac{2C_2^2}{\varepsilon^2 (2 - \varepsilon)^2} \, Mq \, \frac{R}{q} \, R.$$

Since $Mq \le 2\pi/C_1$ due to (4) and $R/q < L$ by assumption, it follows that the first term in (34) is smaller than $\eta/2$ if only R does not exceed some sufficiently small positive number ϱ_1 that depends only on ε, η, L. As

$$\frac{M R^4}{q^2} = Mq \left(\frac{R}{q} \right)^3 R,$$

we see analogously that the second term in (34) becomes smaller than $\eta/2$ provided $R < \varrho_2$. Therefore (29) implies that (30) is true for $R < \varrho_0 := \min(\varrho_1, \varrho_2)$. $\square$

Theorem 4.2. *Let $d \geq 2$, $\varepsilon \in (0,1)$, $\eta \in (0,1)$, $L \in (1,\infty)$, and suppose the set $\mathcal{X}$ has partition norm R and separation distance q. Then there exists a positive number $\varrho_0 = \varrho_0(d,\varepsilon,\eta,L) > 0$ such that*

$$\mathbb{P}\left(1-\varepsilon \leq \frac{\|\boldsymbol{f}\|_{\boldsymbol{W},2}}{\|f\|_2} \leq 1+\varepsilon\right) \geq 1-\eta \tag{35}$$

for every polynomial degree $N \geq 0$ whenever the uniformity condition $R/q < L$ and the density condition $R < \varrho_0$ hold.

Proof. From (8) we infer that

$$|(AA^*)_{jj}| = w_j \sum_{k=1}^{n} |Y_k^d(\boldsymbol{\xi}_j)|^2 = w_j \frac{n}{\mu_d(\mathbb{S}^d)}.$$

Thus, (3) gives

$$\sum_{j=1}^{M} |(AA^*)_{jj}|^2 \leq \frac{n^2}{\mu_d(\mathbb{S}^d)^2} \sum_{j=1}^{M} w_j^2 \leq C_2^2 M R^{2d} \frac{n^2}{\mu_d(\mathbb{S}^d)^2}. \tag{36}$$

Now fix j and let $\ell \neq j$. From (7) we obtain that

$$\begin{aligned} |(AA^*)_{j\ell}| &= \sqrt{w_j w_\ell}\left|\sum_{\kappa=0}^{N}\sum_{i=1}^{\mathcal{H}_d(\kappa)} Y_{\kappa,i}^d(\boldsymbol{\xi}_j)\overline{Y_{\kappa,i}^d(\boldsymbol{\xi}_\ell)}\right| \\ &= \sqrt{w_j w_\ell}\left|\sum_{\kappa=0}^{N} \frac{\mathcal{H}_d(\kappa)}{\mu_d(\mathbb{S}^d)} \mathsf{C}_\kappa^{(d-1)/2}(\boldsymbol{\xi}_j \cdot \boldsymbol{\xi}_\ell)\right|. \end{aligned}$$

By virtue of the inequalities (6), (9), and (3) we get

$$\begin{aligned} |(AA^*)_{j\ell}| &\leq C_4 R^d \sum_{\kappa=0}^{N} \kappa^{d-1}\kappa^{-(d-1)/2} |\sin\varphi_{j\ell}|^{-(d-1)/2} \\ &\leq C_5 R^d N^{(d+1)/2} |\sin\varphi_{j\ell}|^{-(d-1)/2}, \end{aligned}$$

where $\varphi_{j\ell} \in (0,\pi]$ is determined by $\boldsymbol{\xi}_j \cdot \boldsymbol{\xi}_\ell = \cos\varphi_{j\ell}$. Due to symmetry, we only consider the northern hemisphere $\varphi_{j\ell} \in (0,\pi/2]$ for which we denote $I_j = \{\ell \in \mathbb{N} : 1 \leq \ell \leq M, q \leq \varphi_{j\ell} \leq \pi/2\}$. The southern hemisphere can be treated analogously, except for the south pole region itself. For this region, however, we may use estimate (36). In the northern hemisphere we have

$$\sum_{j=1}^{M}\sum_{\ell\in I_j} |(AA^*)_{j\ell}|^2 \leq C_6 R^{2d} N^{d+1} \sum_{j=1}^{M}\sum_{\ell\in I_j} |\sin\varphi_{j\ell}|^{-(d-1)}. \tag{37}$$

For $1 \leq m \leq \pi/(2q)$, put

$$S_m = \{\boldsymbol{\xi} \in \mathbb{S}^d : (m-1)q < d(\boldsymbol{\xi},\boldsymbol{\xi}_j) \leq mq\};$$

in the notation S_m we suppress the dependence on the number j, which was fixed at the very beginning. We denote by M_m the number of sampling nodes $\boldsymbol{\xi}_\ell$ that belong to S_m. These numbers have been estimated in [11] by $M_m \le C_7 m^{d-1}$. Since $(m-1)q < \varphi_{j\ell} \le mq$ for $\boldsymbol{\xi}_\ell \in S_m$, we obtain that

$$\sum_{\ell \in I_j} |\sin \varphi_{j\ell}|^{-(d-1)} \le C_8 \sum_{m=1}^{\lceil \pi/(2q) \rceil} \frac{m^{d-1}}{(mq)^{d-1}} \le C_9 q^{-d}.$$

Inserting this in (37) we get

$$\sum_{j=1}^{M} \sum_{\ell \in I_j} |(AA^*)_{j\ell}|^2 \le C_{10} M R^{2d} N^{d+1} q^{-d}. \tag{38}$$

We now take (36) for the point at the north pole and for the south pole region (which contains at most one point) and (38) for the remaining matrix entries. Addition of these estimates gives

$$\frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{n^2 \varepsilon^2 (2-\varepsilon)^2} \le \frac{C_{11} M R^{2d}}{\varepsilon^2 (2-\varepsilon)^2} + \frac{C_{12} M R^{2d}}{N^{d-1} q^d \varepsilon^2 (2-\varepsilon)^2}. \tag{39}$$

The first term on the right of (39) does not exceed $\eta/2$ if $R < \varrho_1 = \varrho_1(d, \varepsilon, \eta, L)$ because

$$M R^{2d} = M q^d \left(\frac{R}{q}\right)^d R^d < \frac{\mu_d(\mathbb{S}^d)}{C_1} L^d R^d$$

by (4) and by our assumption. If $NR \le \varepsilon / B_d$, then $1 - \varepsilon \le \|f\|_{\boldsymbol{W},2} / \|f\|_2 \le 1 + \varepsilon$ is true deterministically by (11). So let $NR > \varepsilon / B_d$. Then by (4) and by assumption

$$\frac{M R^{2d}}{N^{d-1} q^d} < \frac{B_d^{d-1}}{\varepsilon^{d-1}} \frac{R^{3d-1} M}{q^d} = \frac{B_d^{d-1}}{\varepsilon^{d-1}} M q^d \left(\frac{R}{q}\right)^{3d-1} q^{d-1} \le \frac{B_d^{d-1}}{\varepsilon^{d-1}} \frac{\mu_d(\mathbb{S}^d)}{C_1} L^{3d-1} R^{d-1}.$$

Thus, the second term on the right of (39) is smaller than $\eta/2$ for all $R < \varrho_2 = \varrho_2(d, \varepsilon, \eta, L)$. Using (29) we arrive at the conclusion that (35) is true whenever $R < \varrho_0 := \min(\varrho_1, \varrho_2)$. $\square$

Remark 4.3. The partition norm R is completely determined by $\mathcal{R}$ alone, while the separation distance q depends on $\mathcal{X}$ only. Let

$$D(\boldsymbol{\xi}_j, \delta) := \{\boldsymbol{\xi} \in \mathbb{S}^d : d(\boldsymbol{\xi}, \boldsymbol{\xi}_j) \le \delta\}, \quad r_j = 2 \sup\{\delta > 0 : D(\boldsymbol{\xi}_j, \delta) \subset R_j\},$$

and put $r = \min_j r_j$. Thus, r is the largest number such that each R_j contains a spherical disk of diameter r centered at $\boldsymbol{\xi}_j$. The advantage of r is that it can be determined by sole knowledge of the pairs $(R_j, \boldsymbol{\xi}_j)$, whereas q is a quantity that depends on the location of $\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}_\ell$ for different j and ℓ. Theorems 4.1 and 4.2 remain literally true with q replaced by r. $\square$

Remark 4.4. If $\ell > 0$ is small, one can partition $\mathbb{S}^d$ into $O(\ell^d)$ small regions which are close to d-dimensional cubes of side-length ℓ. Choose the sampling nodes near the centers of these cube-type regions. Then $R \sim \sqrt{d}\,\ell$, $q \sim \ell$, $r \sim \ell$ as $\ell \to 0$. This shows that sampling pairs as required in Theorems 4.1 and 4.2 do really exist if only $L > \sqrt{d}$. □

Remark 4.5. The following deterministic Marcinkiewicz-Zygmund inequality for trigonometric polynomials $f(\boldsymbol{\xi}) = \sum_{\boldsymbol{k}\in[-N,N]^d\cap\mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\boldsymbol{\xi}}$ on the torus $\mathbb{T}^d$ is taken from [7, 1]:

$$(2 - \mathrm{e}^{\pi dNR})\|f\|_2 \le \|\boldsymbol{f}\|_{\boldsymbol{W},2} \le \mathrm{e}^{\pi dNR}\|f\|_2,$$

where $R = \max_j \max_{\boldsymbol{\xi},\boldsymbol{\eta}\in R_j} \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_\infty$ for some partition $\{R_j \subset \mathbb{T}^d\}$. Consequently, for $NR \le \log(1+\varepsilon)/(\pi d)$ we have

$$(1-\varepsilon)\|f\|_2 \le \|\boldsymbol{f}\|_{\boldsymbol{W},2} \le (1+\varepsilon)\|f\|_2.$$

All the probabilistic results established here have analogues on the torus. The probabilistic Marcinkiewicz-Zygmund inequality as given in Theorem 3.1 follows from [3, Theorem 7.1]. In order to show the result of Theorem 4.2 for the torus we have to estimate the Frobenius norm $\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2$ where $\boldsymbol{A}$ is the weighted nonequispaced Fourier matrix $\boldsymbol{A} = \left(\sqrt{w_j}\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\boldsymbol{\xi}_j}\right)_{j=1,\dots,M,\boldsymbol{k}\in[-N,N]^d\cap\mathbb{Z}^d}$. It is easily checked that $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = (2N+1)^d$ and that the off-diagonal decay is governed by

$$|(AA^*)_{j,\ell}| = \sqrt{w_j w_\ell}\left|\sum_{\boldsymbol{k}\in[-N,N]^d\cap\mathbb{Z}^d} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}(\boldsymbol{\xi}_j-\boldsymbol{\xi}_\ell)}\right| \le \begin{cases} w_j\,(2N+1)^d & \text{for} \quad j=\ell, \\ \sqrt{w_j w_\ell}\,\frac{(2N+1)^{d-1}}{2\|\boldsymbol{\xi}_j-\boldsymbol{\xi}_\ell\|_\infty} & \text{for} \quad j\neq\ell. \end{cases}$$

Let $q = \min_{j\neq\ell}\|\boldsymbol{\xi}_j - \boldsymbol{\xi}_\ell\|_\infty$ denote the separation distance of the sampling set. Using the packing argument from [12, Theorem 4.6], we obtain

$$\frac{2\|\boldsymbol{A}\boldsymbol{A}^*\|_{\mathrm{F}}^2}{(2N+1)^{2d}\varepsilon^2(2-\varepsilon)^2} \le \frac{C_{13}MR^{2d}}{\varepsilon^2(2-\varepsilon)^2} + \frac{C_{14}MR^{2d}\psi(q)}{N^2q^d\varepsilon^2(2-\varepsilon)^2}$$

where $\psi(q) = \log(1/q)$ for $d=2$ and $\psi(q) = 1$ for $d \ge 3$. Assuming $NR > \log(1+\varepsilon)/(\pi d)$, we can now proceed as in the proof of Theorem 4.2.

5 Examples

Example 5.1. We consider the two dimensional unit sphere $\mathbb{S}^2$: Choosing $\varepsilon = 1/2$, taking the polynomial degree $N = 13$, and noting that $B_2 \le 132$ we see that if $R \le 1/(2\cdot 13\cdot 132) \approx 2.91\cdot 10^{-4}$, then (11) holds, that is, for all $f \in \Pi_{13}^2$ we have

$$\frac{1}{2}\|f\|_2 \le \|\boldsymbol{f}\|_{\boldsymbol{W},2} \le \frac{3}{2}\|f\|_2.$$

On the other hand, we have $NR \leq 1$ for $N \leq 3432$ and thus Theorem 3.1 yields for randomly chosen $f \in \Pi^2_{3432}$ the inequality

$$\mathbb{P}\left(\frac{1}{2}\|f\|_2 \leq \|\boldsymbol{f}\|_{\boldsymbol{W},2} \leq \frac{3}{2}\|f\|_2\right) \geq 0.99995.$$

Asking for the deterministic result for the degrees $N \leq 3432$ would force us to have $R' \leq 1/(2 \cdot 3432 \cdot 132) \approx 1.10 \cdot 10^{-6}$.

Now, let (θ, φ) with $0 \leq \theta \leq \pi$, $0 \leq \varphi < 2\pi$ be the spherical coordinates on $\mathbb{S}^2$. The $2m$ meridians given by $\varphi = \pi k/m$ ($k = 0, \ldots, 2m-1$) and the $m-1$ parallels of latitude specified by $\theta = \pi k/m$ ($k = 1, \ldots, m-1$) divide $\mathbb{S}^2$ into $2m^2$ regions. Let $\mathcal{R}$ be the set of these $2m^2$ regions and choose exactly one sampling node arbitrarily in each one. Since the regions near the equator look approximately like squares, we have $R \approx \pi\sqrt{2}/m$.

To make things better visible, we pass from $\mathbb{S}^2$ to the surface of the earth, which is assumed to be a sphere of radius 6370 km. Then the sampling nodes that guarantee the deterministic result for $N \leq 13$ or the probabilistic result for $N \leq 3432$ are at an average distance of $\frac{\pi}{m} \cdot 6370$ km $\approx \frac{R}{\sqrt{2}} \cdot 6370$ km ≈ 1.3 km near the equator, whereas the deterministic result for $N \leq 3432$ would force an average equatorial distance of $\frac{\pi}{m'} \cdot 6370\,\text{km} \approx \frac{R'}{\sqrt{2}} \cdot 6370\,\text{km} \approx 5.0\,\text{m}$.

Example 5.2. Divide $\mathbb{S}^3$ into small regions that are close to cubes of side-length ℓ and take exactly one sampling node in each of these regions. This time $R \approx \sqrt{3}\,\ell$ and $M\ell^3 \approx \omega_3 = 2\pi^2$. Let $\varepsilon = 1/2$ and $N = 8$. Since $B_3 \leq 666$, the starting estimate required in (11) is $R \leq 1/(2 \cdot 8 \cdot 666) \approx 9.4 \cdot 10^{-5}$, which is true for $M \approx 2\pi^2/\ell^3 \approx 1.2 \cdot 10^{14}$ sampling nodes. With $\varepsilon = 1/2$ we have Theorem 3.1 for $N \leq 1/R \approx 10656$ and a probability larger than $1 - 10^{-8}$. The deterministic result for $N \leq 10656$ would demand $R' \leq 1/(2 \cdot 10656 \cdot 666) \approx 7.0 \cdot 10^{-8}$.

Assuming that our universe is $r\mathbb{S}^3$ with $r = 18 \cdot 10^9$ light-years, we obtain $rR/\sqrt{3} \approx$ 975000 light-years and $rR'/\sqrt{3} \approx 732$ light-years for the average distances between the sampling nodes. Note that the diameter of our home galaxy is about 100000 light-years.

References

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

[2] A. Böttcher and S. Grudsky. The norm of the product of a large matrix and a random vector. *Electron. J. Probab.*, 8:1 – 29, 2003.

[3] A. Böttcher and D. Potts. Probability against condition number and sampling of multivariate trigonometric random polynomials. *Electron. Trans. Numer. Anal.*, 26:178 – 189, 2007.

[4] F. Filbir and W. Themistoclakis. Polynomial approximation on the sphere using scattered data. *Math. Nachr.*, 281:650 – 668, 2008.

[5] Q. T. Le Gia and H. N. Mhaskar. Localized linear polynomial operators and quadrature formulas on the sphere. *Preprint*, 2007.

[6] L. Golinskii, D. S. Lubinsky, and P. Nevai. Large sieve estimates on arcs of a circle. *J. Number Theory*, 91:206 – 229, 2001.

[7] K. Gröchenig. Reconstruction algorithms in irregular sampling. *Math. Comput.*, 59:181 – 194, 1992.

[8] K. Gröchenig, B. Pötscher, and H. Rauhut. Learning trigonometric polynomials from random samples and exponential inequalities for eigenvalues of random matrices. *Preprint arXiv:math/0701781*, 2007.

[9] K. Jetter, J. Stöckler, and J. D. Ward. Errors estimates for scattered data interpolation on spheres. *Math. Comput.*, 68:733 – 747, 1999.

[10] J. Keiner, S. Kunis, and D. Potts. Efficient reconstruction of functions on the sphere from scattered data. *J. Fourier Anal. Appl.*, 13:435 – 458, 2007.

[11] S. Kunis. A note on stability results for scattered data interpolation on Euclidean spheres. *Adv. Comput. Math.*, 2008, accepted.

[12] S. Kunis and D. Potts. Stability results for scattered data interpolation by trigonometric polynomials. SIAM *J. Sci. Comput.*, 29:1403 – 1419, 2007.

[13] D. S. Lubinsky, A. Máté, and P. Nevai. Quadrature sums involving pth powers of polynomials. SIAM *J. Math. Anal.*, 18:531 – 544, 1987.

[14] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward. Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature. *Math. Comput.*, 70:1113 – 1130, 2001. Corrigendum on the positivity of the quadrature weights in 71:453 – 454, 2002.

[15] C. Müller. *Spherical Harmonics.* Springer, Aachen, 1966.

[16] J. Ortega-Cerda and J. Saludes. Marcinkiewicz-Zygmund inequalities. *J. Approx. Theory*, 145:237 – 252, 2007.

[17] E. Rakhmanov and B. Shekhtman. On discrete norms of polynomials. *J. Approx. Theory*, 139:2–7, 2006.

[18] H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. IEEE *Trans. Inform. Theory*, 2008, accepted.

[19] M. Reimer. *Multivariate Polynomial Approximation.* Birkhäuser, Basel, 2003.

[20] G. Szegő. *Orthogonal Polynomials.* Amer. Math. Soc., Providence, RI, 4th edition, 1975.