A Survey of Sampling Discretization of Integral and Uniform Norms

Feng Dai, Egor D. Kosov, and Vladimir N. Temlyakov*

October 22, 2026

Abstract

This paper surveys recent developments in the sampling discretization of integral and uniform norms for functions in general finite-dimensional spaces. These results generalize the classical Marcinkiewicz–Zygmund inequalities for trigonometric and algebraic polynomials, which play a crucial role in Fourier analysis, interpolation, and approximation theory. We focus on the problem in the broad context of finite-dimensional subspaces, where norms defined by general probability measures are approximated by their discrete counterparts. The primary emphasis is on results closely related to the authors' recent research. A key objective is to highlight the main ideas and techniques that form the foundation of the proofs in this area. This survey serves as a complement to three recently published survey papers on sampling discretization [18, 32, 45].

MSC: 41-02, 41A17, 41A10, 47-02

Keywords: Marcinkiewicz inequality, sampling discretization of integral norms, universal discretization, entropy numbers

1 Introduction 2
2 Connections with other areas 6
3 Marcinkiewicz discretization of L_2 norms 8
3.1 Spectral characterization of weighted discretization . . . 9
3.2 Discretization using random samples 10
3.3 The minimum number of points required for L_2-discretization 14
4 Marcinkiewicz discretization theorems 15
4.1 A conditional theorem 15
4.2 Preliminary discretizations 20
4.3 Sampling discretization of L_p norms 22

*The first named author's research was partially supported by NSERC of Canada Discovery Grant RGPIN-2020-03909. The second named author's research was supported by the AEI grants RYC2023-043616-I and PID2023-150984NB-I00 funded by MICIU/AEI/10.13039/501100011033/ FEDER, EU, and by the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D (CEX2020-001084-M). The second named author thanks CERCA Programme (Generalitat de Catalunya) for institutional support. The third named author's research (Subsections 3.1, 3.2, and Section 7) was supported by the Russian Science Foundation under grant no. 23-71-30001, https://rscf.ru/project/23-71-30001/, and performed at Lomonosov Moscow State University.

Surveys in Approximation Theory
Volume 10, 2026. pp. 1–42.
© 2026 Surveys in Approximation Theory.
ISSN 1555-578X
All rights of reproduction in any form reserved.

4.4 Sampling discretization with weights 24
5 Universal discretization 25
6 Some improved bounds in sampling discretization of integral norms 29
7 Sampling discretization in the uniform norm 32
7.1 General bounds of $m(X_N;\infty;C)$ with C being independent of N 33
7.2 Estimates of $m(X_N;\infty;C_N)$ with C_N being dependent on N 34
References 37

1 Introduction

The classical Marcinkiewicz inequality has the following original form (see [68, Vol. II, p. 30]): For any $1 < p < \infty$, there exists a constant $C_p \geqslant 1$ depending only on p such that for every $f \in \mathcal{T}_n := \operatorname{span}\{e^{\mathbf{i}kx} : k \in \mathbb{Z},\ -n \leqslant k \leqslant n\}$,

$$C_p^{-1}\|f\|_{L_p[0,2\pi]} \leqslant \|f\|_{L_p(\Lambda)} \leqslant C_p\|f\|_{L_p[0,2\pi]}, \tag{1.1}$$

where $\Lambda = \Lambda_{2n+1} := \{\frac{2j\pi}{2n+1} : j = 0, 1, \dots, 2n\}$, and $\|f\|_{L_p(\Lambda)}$ denotes the L_p norm of f defined with respect to the uniform probability measure on Λ; that is,

$$\|f\|_{L_p(\Lambda)} := \begin{cases} \left(\frac{1}{|\Lambda|}\sum\limits_{\omega\in\Lambda}|f(\omega)|^p\right)^{1/p}, & \text{if } 1 \leqslant p < \infty, \\ \max\limits_{\omega\in\Lambda}|f(\omega)|, & \text{if } p = \infty. \end{cases}$$

Throughout this paper, $|E|$ denotes the cardinality of a finite set E. It is important to note that the constant C_p on the right-hand side of (1.1) can be chosen to be independent of p. Indeed, the right-hand inequality in (1.1) remains true at the endpoints $p = 1, \infty$. However, the left-hand inequality requires stronger conditions at the endpoints. It holds for the full range of $1 \leqslant p \leqslant \infty$ with Λ_{4n+1} in place of Λ_{2n+1}.

Historically, the inequality (1.1) for the case $p = \infty$ (discretization of the uniform norm) was first established by Bernstein in 1931-1932 (see [4] and [5]). The result for $1 < p < \infty$ was proved by J. Marcinkiewicz, while A. Zygmund extended the study to the endpoints $p = 1, \infty$ in 1937 (see [32, p.4] and [68, Vol. II, p. 30]). Therefore, inequalities of the form (1.1) are often referred to as Bernstein-type theorems when $p = \infty$, and Marcinkiewicz-type theorems when $1 \leqslant p < \infty$ (see [18, 32, 60, 61]). In broader literature, these results are known as Marcinkiewicz–Zygmund inequalities (see, for instance, [8, 18, 32]). The Marcinkiewcz type inequality (1.1) is a basic tool in classical analysis, playing a crucial role in the study of Fourier series convergence, Lagrange interpolation, sampling discretization, and weighted approximation. For various extensions of these results, we refer the reader to [8, 18, 32, 35, 47, 48, 51] and the references therein.

This survey focuses primarily on recent work by the authors and their collaborators concerning Marcinkiewicz-type and Bernstein-type theorems for general finite-dimensional subspaces. We do not impose structural assumptions on these spaces, except for conditions on the embedding into the space of bounded functions. Our objective is to highlight the key techniques and methods used to establish inequalities of the form (1.1) in this general setting. The primary technical tools rely on concentration inequalities rooted in the work of J. Bourgain, J. Lindenstrauss, V.D. Milman,

G. Schechtman and others (see, e.g., [7, 30, 52] and references thererin), as well as chaining arguments (see [59]). We aim to demonstrate how these techniques apply specifically to the problem of sampling discretization.

It should be noted that this survey is largely confined to results closely related to the authors' recent research on sampling discretization. Given the extensive literature in this area, the references provided here are by no means complete. Our intention is to focus on the results we are most familiar with, rather than covering topics beyond the immediate scope of our work. In particular, we do not discuss in detail the important and related applications in sampling recovery. For these topics, we refer the reader to the surveys [24, 32] and the significant recent contributions by D. Krieg, M. Ullrich, T. Ullrich and their collaborators (see [2, 25, 29, 39, 40, 41, 42, 53, 67] and references therein).

Let us introduce a general framework for the sampling discretization problem we are going to study. Let $(\Omega, \mu, \mathcal{F})$ be a probability space, where $\mathcal{F}$ is a σ-algebra of subsets of Ω and μ is a probability measure on $\mathcal{F}$. Given $1 \leqslant p < \infty$, let $L_p(\Omega, \mu)$ denote the L_p-space defined with respect to the measure μ on Ω, equipped with the norm $\|\cdot\|_{L_p(\Omega,\mu)}$ given by

$$\|f\|_{L_p(\Omega,\mu)} = \Big(\int_\Omega |f|^p \, d\mu\Big)^{1/p}.$$

Let X_N denote a linear space of $\mathcal{F}$-measurable functions on Ω of dimension N (where N is typically large). We assume that every function in X_N is defined everywhere on Ω, and satisfies

$$\|f\|_{L_\infty(\Omega)} := \sup_{x\in\Omega} |f(x)| < \infty.$$

Finally, we denote the unit ball of X_N in $L_p(\Omega, \mu)$ by

$$X_N^p := \big\{f \in X_N \colon \|f\|_{L_p(\Omega,\mu)} \leqslant 1\big\}.$$

The Marcinkiewicz discretization problem with weights. Let $C_2 \geqslant C_1 > 0$ be two given positive constants. The weighted Marcinkiewicz discretization of the L_p norm, $1 \leqslant p < \infty$, in the space X_N approximates the continuous $L_p(\Omega, \mu)$ norm using a discrete signed measure supported on a finite subset $\Omega_m = \{\xi^1, \dots, \xi^m\} \subset \Omega$. Specifically, we require that the integral norm of every $f \in X_N$ is bounded by a weighted discrete sum as follows:

$$C_1 \int_\Omega |f|^p \, d\mu \leqslant \sum_{j=1}^m \lambda_j |f(\xi^j)|^p \leqslant C_2 \int_\Omega |f|^p \, d\mu, \tag{1.2}$$

where the nodes $\xi^j \in \Omega$ and the weights $\lambda_j \in \mathbb{R}$ are independent of $f \in X_N$. If the estimate (1.2) holds for every function f in a subspace X_N, we say that X_N admits a *weighted Marcinkiewicz-type discretization theorem* with parameters p and $m \in \mathbb{N}$, and constants C_1, C_2.

Of particular interest are the following special cases of (1.2) where the weights are uniform, i.e., $\lambda_1 = \cdots = \lambda_m = \dfrac{1}{m}$.

(i) **The Marcinkiewicz discretization problem.** We say that X_N admits a *Marcinkiewicz-type discretization theorem* with parameters p and $m \in \mathbb{N}$, and constants C_1, C_2 if

$$C_1 \|f\|_{L_p(\Omega,\mu)}^p \leqslant \frac{1}{m} \sum_{j=1}^m |f(\xi^j)|^p \leqslant C_2 \|f\|_{L_p(\Omega,\mu)}^p, \quad \forall f \in X_N. \tag{1.3}$$

(ii) **The Marcinkiewicz discretization problem with $\epsilon \in (0,1)$.** We say that X_N admits a *Marcinkiewicz-type discretization theorem with* ϵ and parameters p and $m \in \mathbb{N}$ if

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in X_N. \tag{1.4}$$

The Bernstein discretization problem. The above definition can also be modified to include the case $p = \infty$, where we ask for

$$C_1\|f\|_{L_\infty(\Omega)} \leqslant \max_{1\leqslant j\leqslant m} \left|f\left(\xi^j\right)\right| \leqslant \|f\|_{L_\infty(\Omega)}, \quad \forall f \in X_N.$$

Universal Marcinkiewicz-type discretization problem. We can also consider a universal Marcinkiewicz-type discretization problem for a finite collection $\mathcal{X} := \{X_k\}_{k=1}^L$ of finite-dimensional spaces of functions on Ω. Given $1 \leqslant p < \infty$, we say that a finite set $\{\xi^j\}_{j=1}^m \subset \Omega$ provides *universal discretization* of the L_p norm for the collection $\mathcal{X}$ with positive constants C_1, C_2 if

$$C_1\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant C_2\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in \bigcup_{k=1}^L X_k.$$

Universal weighted discretization can be defined likewise.

In the framework described above, X_N can be either a real or complex space. For Marcinkiewicz discretization with general constants $C_1, C_2 > 0$, results for complex spaces can be derived from the real case by separating the real and imaginary parts. However, this reduction fails for the Marcinkiewicz discretization problem with $\epsilon > 0$, as the separation argument distorts the isometric constants $(1\pm\epsilon)$. For simplicity, we generally assume that X_N is a real space unless stated otherwise. Nevertheless, most results extend to complex spaces using similar techniques.

This survey presents tools and techniques for the Marcinkiewicz discretization problem on finite-dimensional spaces subject to certain additional conditions. Typically, two main types of conditions are used: (i) bounds on entropy numbers; and (ii) Nikol'skii-type inequalities. We postpone the definition and discussion of entropy numbers to Section 4. In this section, we focus on Nikol'skii-type inequalities. These inequalities constitute the main assumption imposed on the subspaces in our setting and appear throughout the paper in the formulation of various results.

Nikol'skii-type inequality. Given $p \in [1,\infty)$, an N-dimensional subspace $X_N \subset L_\infty(\Omega)$ is said to satisfy the (p,∞)-Nikol'skii inequality with constant $M > 0$ if

$$\|f\|_{L_\infty(\Omega)} \leqslant M\|f\|_{L_p(\Omega,\mu)}, \quad \forall f \in X_N, \tag{1.5}$$

in which case we write $X_N \in \mathrm{NI}_{p,\infty}(M)$. Of particular interest is the case of $p = 2$, where the exact constant M in (1.5) can be written explicitly as follows. Let $\{u_j(x)\}_{j=1}^N$ be an orthonormal basis of $X_N \subset L_2(\Omega,\mu)$. Then we have

$$\sup_{f\in X_N, \|f\|_{L_2(\Omega,\mu)}\leqslant 1} |f(x)| = \Big(\sum_{j=1}^N |u_j(x)|^2\Big)^{1/2} =: \sqrt{F(x)}, \quad \forall x \in \Omega, \tag{1.6}$$

which implies

$$X_N \in \mathrm{NI}_{2,\infty}(M) \iff \|F\|_{L_\infty(\Omega)} \leqslant M^2. \tag{1.7}$$

The function

$$\lambda(X_N, \mu; x) := \frac{1}{F(x)},$$

known as the Christoffel function of the space X_N, is independent of the selection of orthonormal basis of X_N. Finally, since

$$\|f\|_{L_2(\Omega,\mu)}^2 \leqslant \|f\|_{L_p(\Omega,\mu)}^p \, \|f\|_{L_\infty(\Omega)}^{2-p}$$

for $1 \leqslant p \leqslant 2$, the condition

$$\|f\|_{L_\infty(\Omega)} \leqslant \sqrt{M}\, \|f\|_{L_2(\Omega,\mu)}$$

implies

$$\|f\|_{L_2(\Omega,\mu)} \leqslant M^{\frac{1}{p}-\frac{1}{2}} \|f\|_{L_p(\Omega,\mu)} \quad \text{and} \quad \|f\|_{L_\infty(\Omega)} \leqslant M^{\frac{1}{p}} \|f\|_{L_p(\Omega,\mu)}. \tag{1.8}$$

In terms of the Nikol'skii inequalities, this establishes the implication

$$X_N \in \mathrm{NI}_{2,\infty}(M^{1/2}) \implies X_N \in \mathrm{NI}_{p,\infty}(M^{1/p}) \ \text{ for any } 1 \leqslant p \leqslant 2. \tag{1.9}$$

Notation. To conclude this section, we introduce some notation that will be used throughout this paper. Given a finite set of points $\Omega_m = \{\xi^1, \dots, \xi^m\}$ and $1 \leqslant p \leqslant \infty$, we denote by $L_p(\Omega_m)$ the space of all functions on Ω_m equipped with the norm

$$\|f\|_{L_p(\Omega_m)} := \begin{cases} \Big(\dfrac{1}{m} \displaystyle\sum_{j=1}^m |f(\xi^j)|^p\Big)^{1/p}, & \text{if } 1 \leqslant p < \infty, \\ \\ \displaystyle\max_{1\leqslant j\leqslant m} |f(\xi^j)|, & \text{if } p = \infty. \end{cases}$$

We identify vectors in $\mathbb{R}^m$ with functions defined on the set

$$\Omega_m := \{1, \dots, m\}.$$

In this special case, we adopt the following conventions for vectors $f \in \mathbb{R}^m$. We denote the normalized norm $\|f\|_{L_p(\Omega_m)}$ by $\|f\|_{L_p^m}$ and the unnormalized norm by

$$\|f\|_{\ell_p^m} := m^{1/p} \|f\|_{L_p^m} = \begin{cases} \Big(\displaystyle\sum_{j=1}^m |f(j)|^p\Big)^{1/p}, & \text{if } 1 \leqslant p < \infty, \\ \\ \displaystyle\max_{1\leqslant j\leqslant m} |f(j)|, & \text{if } p = \infty. \end{cases}$$

Accordingly, we define the unit ball

$$B_{\ell_p^m} := \big\{f \in \mathbb{R}^m \colon \|f\|_{\ell_p^m} \leqslant 1\big\}.$$

More generally, for a normed space $(X, \|\cdot\|_X)$, we define

$$B_X := \big\{f \in X \colon \|f\|_X \leqslant 1\big\}.$$

We also denote the Euclidean norm of $x \in \mathbb{R}^m$ as $|x|$. Unless otherwise stated, the symbols $C, c, C_1, c_1, \dots$ denote positive universal constants.

Organization of the paper. The remainder of the paper is organized as follows.

In Section 2, we discuss the close connections between the Marcinkiewicz discretization problem and several related areas, including embeddings of finite-dimensional subspaces, frames, spectral properties, and operator norms of submatrices of a given matrix.

In Section 3, we survey recent results on Marcinkiewicz sampling discretization for L_2 norms, which is the most extensively studied case.

In Section 4, we present one of the general technical approaches to the Marcinkiewicz sampling discretization of L_p norms for all $1 \leqslant p < \infty$. It is stated in the form of the conditional sampling discretization theorem, which highlights the connection between the Marcinkiewicz discretization problem and entropy numbers of compact subsets of bounded functions under the uniform metric. We also outline the key ideas used in the proofs.

In Section 5, we discuss universal discretization, and its connection with the Restricted Isometry Property, a fundamental concept in compressed sensing.

Section 6 covers recent improvements in bounds for sampling discretization of integral norms, along with the main ideas behind their proofs.

Finally, in Section 7, we briefly discuss some recent progress on sampling discretization in the uniform norm.

2 Connections with other areas

The Marcinkiewicz discretization problem is closely connected to several other areas. In this section, we provide a brief overview of these interactions and reformulate the problem in different contexts. We refer to the survey paper [32] for a detailed discussion of these connections.

First, there are deep results on embeddings of finite-dimensional subspaces of $L_p(\Omega,\mu)$ into ℓ_p^N, which are closely related to the sampling discretization of the L_p norms (see, for instance, [7, 30]). Techniques developed for these embedding problems have proven highly effective for sampling discretization (see [30]). A detailed discussion of this connection is available in Section 4.2 of [32].

Second, there is a close connection between the Marcinkiewicz discretization of L_p norms and the concept of (Λ,p)-frames (see Section 5 of [14]). We recall the following definition from [14, Definition 5.2].

Definition 2.1 ([14])**.** Let $\Lambda := \{\lambda_j\}_{j=1}^m \subset \mathbb{R}$ and $1 \leqslant p < \infty$. A sequence $\{\psi_j\}_{j=1}^m$ of functions in X_N is called a (Λ,p)-frame of X_N with positive constants A and B if for any $f \in X_N$ we have

$$A\|f\|_{L_p(\Omega,\mu)}^p \leqslant \sum_{j=1}^m \lambda_j |\langle f,\psi_j\rangle|^p \leqslant B\|f\|_{L_p(\Omega,\mu)}^p,$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of $L_2(\Omega,\mu)$. When $\lambda_1 = \cdots = \lambda_m = 1$, a (Λ,p)-frame is called a p-frame (see [1]).

The weights in Definition 2.1 may have either sign, as in the general weighted discretization inequality (1.2). Nonnegative weights form a special case, which will be the focus of Section 3.

Let $\mathcal{D}(X_N,\cdot,\cdot)$ denote the reproducing kernel (Dirichlet kernel) of the space $X_N \subset L_2(\Omega,\mu) \cap L_p(\Omega,\mu)$; namely,

$$\mathcal{D}(X_N,x,y) := \sum_{j=1}^N u_j(x)u_j(y), \quad x,y \in \Omega$$

where $\{u_j\}_{j=1}^N$ is an orthonormal basis of $(X_N, \|\cdot\|_{L_2(\Omega,\mu)})$. Note that $\mathcal{D}(X_N,\cdot,\cdot)$ is independent of the choice of the orthonormal basis. The Marcinkiewicz inequality (1.2) is equivalent to

$$C_1\|f\|^p_{L_p(\Omega,\mu)} \leqslant \sum_{j=1}^m \lambda_j |\langle f, \mathcal{D}(X_N,\xi^j,\cdot)\rangle|^p \leqslant C_2\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in X_N.$$

In other words, in terms of (Λ, p)-frames, X_N admits a weighted Marcinkiewicz-type discretization theorem of the form (1.2) with constants $C_1, C_2 > 0$ if and only if there exist points $\xi^1,\dots,\xi^m \in \Omega$ and weights $\Lambda = \{\lambda_j\}_{j=1}^m \subset \mathbb{R}$ such that the sequence $\{\mathcal{D}(X_N,\xi^j,\cdot)\}_{j=1}^m$ forms a (Λ,p)-frame of X_N with constants $C_1, C_2 > 0$.

Third, there is a probabilistic approach to the Marcinkiewicz discretization problem with $\epsilon \in (0,1)$. It aims to establish that for IID random points $\xi^1,\dots,\xi^m$ drawn from μ on Ω, and $1 \leqslant p < \infty$, the inequality

$$\sup_{f\in X_N^p} \Big| \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p - \int_\Omega |f|^p \, d\mu \Big| < \epsilon \tag{2.1}$$

holds with high probability. In particular, this implies that X_N admits a Marcinkiewicz-type discretization theorem of the form (1.4). This approach connects to moment estimates of random vectors in probability. Let ξ be a random point drawn from μ. Given a basis $\{u_k\}_{k=1}^N$ of X_N, consider the vector-valued function $\mathbf{u} = (u_1,\dots,u_N)$ on Ω, and the random vector $\boldsymbol{\eta} = \mathbf{u}(\xi)$. Let $\boldsymbol{\eta}^j = \mathbf{u}(\xi^j)$, $j = 1,\dots,m$, be independent copies of $\boldsymbol{\eta}$. Since each function $f \in X_N$ has a unique representation $f = \langle \mathbf{y}, \mathbf{u}\rangle$ with $\mathbf{y} = (y_1,\dots,y_N) \in \mathbb{R}^N$, the inequality (2.1) above is equivalent to

$$\sup_{\mathbf{y}\in K} \Big| \frac{1}{m}\sum_{j=1}^m |\langle \mathbf{y}, \boldsymbol{\eta}^j\rangle|^p - \mathbb{E}|\langle \mathbf{y}, \boldsymbol{\eta}\rangle|^p \Big| < \epsilon, \tag{2.2}$$

where $K := \{\mathbf{y} \in \mathbb{R}^N : \mathbb{E}|\langle \mathbf{y}, \boldsymbol{\eta}\rangle|^p \leqslant 1\}$ is a symmetric convex set in $\mathbb{R}^N$ and $\langle \mathbf{x}, \mathbf{y}\rangle$ denotes the dot product $\sum_{j=1}^N x_j y_j$ of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$. Thus, the probabilistic approach to the Marcinkiewicz discretization problem asks: How many independent copies of a random vector $\boldsymbol{\eta}$ are needed to ensure that (2.2) holds with high probability?

Fourth, the Marcinkiewicz discretization in finite-dimensional spaces relates to the spectral properties and operator norms of submatrices of a given matrix. It can be reduced to a corresponding matrix problem. As above, we fix a basis $\{u_k\}_{k=1}^N$ of X_N, and define the vector-valued function $\mathbf{u} = (u_1,\dots,u_N)$ on Ω. Given a set $\boldsymbol{\xi} = \{\xi^1, \xi^2, \dots, \xi^m\}$ of distinct points in Ω, consider the $m \times N$ matrix

$$\boldsymbol{\Phi}(\boldsymbol{\xi}) := \begin{bmatrix} \mathbf{u}(\xi^1) \\ \mathbf{u}(\xi^2) \\ \vdots \\ \mathbf{u}(\xi^m) \end{bmatrix} = \begin{bmatrix} u_1(\xi^1) & u_2(\xi^1) & \cdots & u_N(\xi^1) \\ u_1(\xi^2) & u_2(\xi^2) & \cdots & u_N(\xi^2) \\ \vdots & \vdots & \vdots & \vdots \\ u_1(\xi^m) & u_2(\xi^m) & \cdots & u_N(\xi^m) \end{bmatrix}.$$

Given $1 \leqslant p \leqslant \infty$, define the following norm on $\mathbb{R}^N$:

$$\|\mathbf{y}\|_{L_p} = \|\langle \mathbf{y}, \mathbf{u}\rangle\|_{L_p(\Omega,\mu)}, \quad \mathbf{y} = (y_1,\dots,y_N) \in \mathbb{R}^N.$$

The Marcinkiewicz inequality (1.3) can then be reformulated as

$$C_1\|\mathbf{y}\|^p_{L_p} \leqslant \big\|\boldsymbol{\Phi}(\boldsymbol{\xi})\mathbf{y}\big\|^p_{L_p^m} \leqslant C_2\|\mathbf{y}\|^p_{L_p}, \quad \forall \mathbf{y} \in \mathbb{R}^N.$$

If $p = 2$ and $\{u_1, \dots, u_N\}$ is an orthonormal basis of $(X_N, \|\cdot\|_{L_2(\Omega,\mu)})$, this is equivalent to

$$C_1\|\mathbf{y}\|_{\ell_2^m}^2 \leqslant \left\|\mathbf{\Phi}(\boldsymbol{\xi})\mathbf{y}\right\|_{L_2^m}^2 \leqslant C_2\|\mathbf{y}\|_{\ell_2^m}^2, \quad \forall \mathbf{y} \in \mathbb{R}^N,$$

which can be reformulated in terms of the spectrum of the sampling matrix $\mathbf{\Phi}(\boldsymbol{\xi})$. Specifically, this translates into the following estimates for the maximum and minimum singular values:

$$\sigma_{\max}(\mathbf{\Phi}(\boldsymbol{\xi})) \leqslant \sqrt{C_2 m} \ \text{ and } \ \sigma_{\min}(\mathbf{\Phi}(\boldsymbol{\xi})) \geqslant \sqrt{C_1 m}.$$

If $\Omega = \Omega_M = \{x_1, \dots, x_M\}$ is a finite set, then the vector-valued function $\mathbf{u}$ can be represented as a matrix with M rows and N columns:

$$\mathbf{\Phi} := \begin{bmatrix} \mathbf{u}(x_1) \\ \mathbf{u}(x_2) \\ \vdots \\ \mathbf{u}(x_M) \end{bmatrix} = \begin{bmatrix} u_1(x_1) & u_2(x_1) & \cdots & u_N(x_1) \\ u_1(x_2) & u_2(x_2) & \cdots & u_N(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ u_1(x_M) & u_2(x_M) & \cdots & u_N(x_M) \end{bmatrix}.$$

In this case, we seek an $m \times N$ submatrix of the full matrix $\mathbf{\Phi}$, denoted by

$$\mathbf{\Phi}(\boldsymbol{\xi}) := \begin{bmatrix} \mathbf{u}(\xi^1) \\ \mathbf{u}(\xi^2) \\ \vdots \\ \mathbf{u}(\xi^m) \end{bmatrix}, \quad \boldsymbol{\xi} = (\xi^1, \dots, \xi^m) \in \Omega_M^m,$$

such that for every $\mathbf{y} = (y_1, \dots, y_N)^T \in \mathbb{R}^N$,

$$C_1\|\mathbf{\Phi}\mathbf{y}\|_{L_p(\Omega_M)}^p \leqslant \|\mathbf{\Phi}(\boldsymbol{\xi})\mathbf{y}\|_{L_p^m}^p \leqslant C_2\|\mathbf{\Phi}\mathbf{y}\|_{L_p(\Omega_M)}^p.$$

Here, setting $f_{\mathbf{y}} := \sum_{j=1}^N y_j u_j$, we have

$$\|\mathbf{\Phi}\mathbf{y}\|_{L_p(\Omega_M)}^p = \frac{1}{M}\sum_{j=1}^M |f_{\mathbf{y}}(x_j)|^p \ \text{ and } \ \|\mathbf{\Phi}(\boldsymbol{\xi})\mathbf{y}\|_{L_p^m}^p = \frac{1}{m}\sum_{j=1}^m |f_{\mathbf{y}}(\xi^j)|^p.$$

Finally, the universal Marcinkiewicz-type discretization problem can alternatively be reformulated in terms of certain properties of matrices. In particular, the universal discretization of the L_2 norm is closely linked with the concept of the Restricted Isometry Property (RIP), which plays significant roles in compressed sensing. A detailed discussion of this connection will be presented in Section 5 of this paper.

3 Marcinkiewicz discretization of L_2 norms

In this section, we survey some known results on the Marcinkiewicz sampling discretization of the L_2 norms, which is the most studied case. Let X_N be an N-dimensional subspace of $L_2(\Omega, \mu)$. For simplicity, we assume all functions in X_N are real-valued. Our goal is to discretize the L_2 norm in X_N via a finite discrete sum:

$$(1-\epsilon)\int_\Omega |f|^2\, d\mu \leqslant \sum_{j=1}^m \lambda_j |f(\xi^j)|^2 \leqslant (1+\epsilon)\int_\Omega |f|^2\, d\mu, \quad \forall f \in X_N, \tag{3.1}$$

where $\xi^j \in \Omega$, $\lambda_j \geqslant 0$, $j = 1, 2, \dots, m$, and $\epsilon \in (0, 1)$. Here we restrict attention to nonnegative weights; zero weights may be omitted. The spectral characterization in Lemma 3.1 below also holds for arbitrary real weights.

3.1 Spectral characterization of weighted discretization

The discretization (3.1) of the L_2 norms can be equivalently formulated using spectral norms of certain symmetric matrices. To this end, we first recall some useful results from matrix analysis.

Let A be an $N \times N$ real symmetric matrix. By the spectral theorem, one can diagonalize A using a sequence of N real eigenvalues,

$$\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_N(A),$$

and an orthonormal basis of eigenvectors $u_1(A), \ldots, u_N(A) \in \mathbb{R}^N$ such that

$$A\big(u_i(A)\big) = \lambda_i(A) u_i(A), \quad i = 1, 2, \ldots, N.$$

Let $\lambda_{\max}(A) := \lambda_1(A)$ and $\lambda_{\min}(A) := \lambda_N(A)$. Then we have

$$\lambda_{\max}(A) = \max_{x \in \mathbb{S}^{N-1}} \langle Ax, x\rangle \quad \text{and} \quad \lambda_{\min}(A) = \min_{x \in \mathbb{S}^{N-1}} \langle Ax, x\rangle, \tag{3.2}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in $\mathbb{R}^N$. In general, the Courant–Fischer min-max theorem asserts that for each $1 \leqslant k \leqslant N$,

$$\lambda_k(A) = \max_{\dim V = k} \min_{x \in V, |x|=1} \langle Ax, x\rangle$$

and

$$\lambda_k(A) = \min_{\dim V = N-k+1} \max_{x \in V, |x|=1} \langle Ax, x\rangle,$$

where V ranges over all subspaces of $\mathbb{R}^N$ with the indicated dimension. If A is positive semidefinite, $\lambda_{\max}(A)$ is the spectral norm of A, which coincides with the operator norm $\|A\|$ of A:

$$\|A\| := \max_{x \in \mathbb{S}^{N-1}} |Ax|.$$

For a general symmetric matrix, we have

$$\|A\| = \max_{x \in \mathbb{S}^{N-1}} |\langle Ax, x\rangle| = \max\{|\lambda_{\max}(A)|, |\lambda_{\min}(A)|\}.$$

Now we give an alternative formulation of (3.1) using spectral norms. Let $\{u_1, \ldots, u_N\}$ be an orthonormal basis of $X_N \subset L_2(\Omega, \mu)$, and consider the vector-valued function

$$\mathbf{u}(\xi) := (u_1(\xi), \ldots, u_N(\xi)), \quad \xi \in \Omega.$$

Define

$$\mathbf{G}(\xi) := \mathbf{u}(\xi) \otimes \mathbf{u}(\xi) = \Big[u_i(\xi) u_j(\xi)\Big]_{1 \leqslant i,j \leqslant N} \in \mathbb{R}^{N \times N}, \quad \xi \in \Omega.$$

Then for each $f = \sum_{j=1}^N a_j u_j = \langle \mathbf{u}, \mathbf{a}\rangle \in X_N$ with $\mathbf{a} = (a_1, \ldots, a_N)^T \in \mathbb{R}^N$, we have

$$\sum_{j=1}^m \lambda_j |f(\xi^j)|^2 - \|f\|^2_{L^2(\mu)} = \mathbf{a}^T \Big(\sum_{j=1}^m \lambda_j \mathbf{G}(\xi^j) - I_N \Big) \mathbf{a}.$$

Here we treat $\mathbf{a}$ as a column vector and $\mathbf{u}(x)$ as a row vector. This together with (3.2) yields the following alternative formulation of (3.1).

Lemma 3.1. *Let $\xi^1, \dots, \xi^m \in \Omega$, $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ and $\epsilon \in (0,1)$. Then the inequalities in* (3.1) *hold for every $f \in X_N$ if and only if*

$$\Big\| I_N - \sum_{j=1}^{m} \lambda_j \mathbf{G}(\xi^j) \Big\| \leqslant \epsilon,$$

where I_N denotes the $N \times N$ identity matrix, and $\|\cdot\|$ denotes the spectral norm.

Lemma 3.1 establishes a connection between the L_2-discretization (3.1) and the following nonlinear m-term approximation problem:

$$\sigma_m(I_N, \mathcal{D}) = \inf_{B \in \Sigma_m(\mathcal{D})} \|I_N - B\|,$$

where $\mathcal{D} := \{\mathbf{G}(\xi) \colon \xi \in \Omega\}$ and $\Sigma_m(\mathcal{D})$ denotes the set of all $N \times N$ matrices of the form

$$B = \sum_{j=1}^{m} \lambda_j \mathbf{G}(\xi^j), \quad \lambda_1, \dots, \lambda_m \in \mathbb{R}, \quad \xi^1, \dots, \xi^m \in \Omega.$$

Using this connection, one can apply a relaxed greedy algorithm to construct a set of points $\{\xi^j\}_{j=1}^m \subset \Omega$, with m of order N^2, for which (3.1) holds with uniform weights $\lambda_1 = \cdots = \lambda_m = \frac{1}{m}$ (see [61, Proposition 5.1]).

3.2 Discretization using random samples

We start with the following result, which was derived in [61] using Lemma 3.4 below.

Theorem 3.2 ([61])**.** *Let X_N be an N-dimensional space of bounded functions on Ω such that $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ for some constant $K \geqslant 1$. Let $\xi^1, \dots, \xi^m$ be a sequence of IID random points drawn from the probability measure μ on Ω. If $m \geqslant 3AKN\epsilon^{-2}\log N$ for some constants $A > 1$ and $\epsilon \in (0,1)$, then the inequality*

$$(1-\epsilon)\|f\|^2_{L_2(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^2 \leqslant (1+\epsilon)\|f\|^2_{L_2(\Omega,\mu)} \tag{3.3}$$

holds simultaneously for all $f \in X_N$ with probability at least $1 - 2N^{-A+1}$.

Remark 3.3. By setting $A = \frac{m\epsilon^2}{3KN\log N}$, we may reformulate the conclusion of Theorem 3.2 as follows. If $m \geqslant 3KN\epsilon^{-2}\log N + 1$, then the estimate (3.3) holds simultaneously for all $f \in X_N$ with probability at least

$$1 - 2N\exp\Big(-\frac{m\varepsilon^2}{3KN}\Big).$$

We note that results of this type can be traced back to Rudelson [56], whose work implies a similar discretization bound for random points but yields a slightly weaker probability estimate.

The proof of Theorem 3.2 relies on the following result of Tropp [66].

Lemma 3.4 ([66])**.** *Let $\{T_k\}_{k=1}^m$ be a sequence of independent $N \times N$ positive semidefinite random matrices, and define $T = \sum_{k=1}^m T_k$. Assume that*

$$\max_{1 \leqslant k \leqslant m} \|T_k\| \leqslant M < \infty \quad \textit{almost surely}$$

for some constant $M > 0$, where $\|\cdot\|$ denotes the spectral norm. Let $\lambda_{\min} := \lambda_{\min}(\mathbb{E}T)$ and $\lambda_{\max} := \lambda_{\max}(\mathbb{E}T)$. Then

$$\mathbb{P}\Big[\lambda_{\min}(T) \leqslant (1-\epsilon)\lambda_{\min}\Big] \leqslant N\left(\frac{1}{e^{\epsilon}(1-\epsilon)^{1-\epsilon}}\right)^{\lambda_{\min}/M}, \quad \forall \epsilon \in [0,1),$$
$$\mathbb{P}\Big[\lambda_{\max}(T) \geqslant (1+\epsilon)\lambda_{\max}\Big] \leqslant N\left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\lambda_{\max}/M}, \quad \forall \epsilon > 0.$$

Remark 3.5. Let $\varphi(t) := (1-t)\log(1-t)$ for $t \in [0,1)$. Since

$$\varphi'''(t) = \frac{1}{(1-t)^2} > 0, \quad \forall t \in [0,1),$$

it follows by Taylor's theorem that

$$\varphi(t) \geqslant \varphi(0) + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2 = -t + \frac{t^2}{2}, \quad \forall t \in [0,1),$$

which implies

$$(1-t)^{1-t} \geqslant \exp\Big(-t + \frac{t^2}{2}\Big), \quad \forall t \in [0,1].$$

Similarly, applying Taylor's theorem to the function $\psi(t) := (1+t)\log(1+t)$ at $t = 0$, and using the fact that $\psi''(t) = \frac{1}{1+t}$ is a decreasing function of $t \geqslant 0$, one has

$$(1+t)\log(1+t) \geqslant t + \frac{1}{2}\frac{t^2}{1+t}, \quad \forall t \geqslant 0,$$

which implies

$$(1+t)^{1+t} \geqslant \exp\Big(t + \frac{1}{2}\frac{t^2}{1+t}\Big), \quad \forall t \geqslant 0.$$

Substituting these estimates into Lemma 3.4 yields

$$\mathbb{P}\Big[\lambda_{\min}(T) \leqslant (1-\epsilon)\lambda_{\min}\Big] \leqslant N\exp\Big(-\frac{\epsilon^2}{2}\frac{\lambda_{\min}}{M}\Big), \quad \forall \epsilon \in [0,1), \tag{3.4}$$
$$\mathbb{P}\Big[\lambda_{\max}(T) \geqslant (1+\epsilon)\lambda_{\max}\Big] \leqslant N\exp\Big(-\frac{\epsilon^2}{2(1+\epsilon)}\frac{\lambda_{\max}}{M}\Big), \quad \forall \epsilon > 0. \tag{3.5}$$

For completeness, we provide a proof of Theorem 3.2 as follows.

Proof of Theorem 3.2. Let $\{u_1, \ldots, u_N\}$ be an orthonormal basis of X_N, and let

$$\mathbf{u}(x) := (u_1(x), \ldots, u_N(x)) \text{ for } x \in \Omega.$$

The condition $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ implies (see (1.7))

$$|\mathbf{u}(x)|^2 = \sum_{j=1}^{N} |u_j(x)|^2 \leqslant KN, \quad \forall x \in \Omega.$$

Now consider the $m \times N$ random matrix

$$\mathbf{\Phi} u = \begin{bmatrix} \mathbf{u}(\xi^1) \\ \mathbf{u}(\xi^2) \\ \vdots \\ \mathbf{u}(\xi^m) \end{bmatrix}.$$

Let $T := \frac{1}{m}\mathbf{\Phi}^T\mathbf{\Phi}$. Then

$$T = \frac{1}{m}\sum_{j=1}^{m} \mathbf{u}(\xi^j)^T\mathbf{u}(\xi^j) = \sum_{j=1}^{m} T_j \quad \text{with} \quad T_j = \frac{1}{m}\mathbf{u}(\xi^j)^T\mathbf{u}(\xi^j).$$

Clearly, the T_j are independent, positive semidefinite random matrices satisfying

$$\|T_j\| = \max_{x\in\mathbb{S}^{N-1}} x^T T_j x = \frac{1}{m}|\mathbf{u}(\xi^j)|^2 \leqslant \frac{KN}{m}.$$

Moreover, since $\{u_j\}_{j=1}^N$ is an orthonormal basis of X_N, we have

$$\mathbb{E}T = \frac{1}{m}\sum_{j=1}^{m} \mathbb{E}\Big[\mathbf{u}(\xi^j)^T\mathbf{u}(\xi^j)\Big] = \mathbb{E}\Big[\mathbf{u}(\xi^1)^T\mathbf{u}(\xi^1)\Big] = I_N,$$

where I_N denotes the $N \times N$ identity matrix. Now applying Lemma 3.4, (3.4) and (3.5) with $M = \frac{KN}{m}$ and $\lambda_{\max} = \lambda_{\min} = 1$ yields that for any $\epsilon \in (0, 1/2)$,

$$\begin{aligned}
&\mathbb{P}\Big[1-\epsilon \leqslant \lambda_{\min}(T) \leqslant \lambda_{\max}(T) \leqslant 1+\epsilon\Big] \\
&\geqslant 1 - \mathbb{P}\Big[\lambda_{\min}(T) \leqslant 1-\epsilon\Big] - \mathbb{P}\Big[\lambda_{\max}(T) \geqslant 1+\epsilon\Big] \\
&\geqslant 1 - N\left[\exp\Big(-\frac{\epsilon^2}{2}\frac{m}{KN}\Big) + \exp\Big(-\frac{\epsilon^2}{2(1+\epsilon)}\frac{m}{KN}\Big)\right] \\
&\geqslant 1 - 2N\exp\Big(-\frac{m\epsilon^2}{3KN}\Big) \geqslant 1 - 2N^{-A+1}
\end{aligned}$$

provided that $m \geqslant 3KAN\epsilon^{-2}\log N$. Theorem 3.2 then follows by Lemma 3.1. □

Remark 3.6. Clearly, the above proof also yields the following estimate: for any $\epsilon \geqslant 1$ and any $m \geqslant 1$,

$$\begin{aligned}
\mathbb{P}\Big[\lambda_{\max}(T) \leqslant 1+\epsilon\Big] &\geqslant 1 - N\exp\Big(-\frac{\epsilon^2}{2(1+\epsilon)}\frac{m}{KN}\Big) \\
&\geqslant 1 - N\exp\Big(-\frac{\epsilon}{4}\frac{m}{KN}\Big).
\end{aligned}$$

In particular, if $1 \leqslant m \leqslant KN\log N$, then taking $\epsilon = \frac{4AKN\log N}{m}$ for some constant $A > 1$, we deduce

$$\mathbb{P}\Big[\lambda_{\max}(T) \leqslant 1 + \frac{4AKN\log N}{m}\Big] \geqslant 1 - N^{-A+1}.$$

Since for any $\mathbf{y} \in \mathbb{R}^N$,

$$\langle T\mathbf{y}, \mathbf{y}\rangle = \frac{1}{m}\sum_{j=1}^{m} |\langle \mathbf{y}, \mathbf{u}(\xi^j)\rangle|^2 \leqslant \lambda_{\max}(T)\cdot |y|^2,$$

this implies the following one-sided L_2-Marcinkiewicz inequality.

Corollary 3.7. *Let X_N be an N-dimensional space of bounded functions on Ω, and assume that $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ for some constant $K \geqslant 1$. Let $\xi^1, \dots, \xi^m$ be a sequence of IID random points drawn from the probability measure μ on Ω. If $1 \leqslant m \leqslant KN\log N$, then for any given constant $A>1$, the one-sided inequality*

$$\frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^2 \leqslant \Big(1 + \frac{4AKN\log N}{m}\Big)\|f\|^2_{L_2(\Omega,\mu)}$$

holds simultaneously for all $f \in X_N$ with probability at least $1 - N^{-A+1}$.

The Nikol'skii condition $\mathrm{NI}_{2,\infty}(\sqrt{KN})$ plays an important role in the above proof of Theorem 3.2. However, the constant K in this assumption can be very large, in which case one cannot expect a Marcinkiewicz-type theorem with uniform weights and at most $CN\log N$ points with $C > 0$ being a universal constant. For instance, consider the space $\mathcal{P}_N$ of all algebraic polynomials of degree at most $N-1$ on $[-1,1]$. It is well known that $\mathcal{P}_N$ satisfies the $(2,\infty)$-Nikol'skii condition $\mathrm{NI}_{2,\infty}(\sqrt{KN})$ with a least constant $K \geqslant CN$, and any Marcinkiewicz-type inequality with uniform weights for the space $\mathcal{P}_N$ requires at least CN^2 points (see [12, Remark 2.1]).

Nonetheless, Theorem 3.2 with $K = 1$ combined with a change of density argument could be used to yield a weighted discretization result for spaces X_N that may not satisfy the $\mathrm{NI}_{2,\infty}(\sqrt{KN})$ assumption or satisfy it only with a very large constant K. To be more precise, let $\lambda(X_N, \mu; \cdot)$ denote the Christoffel function on Ω associated with the space X_N and the measure μ, defined as

$$\lambda(X_N, \mu; x) := \inf_{\substack{f\in X_N \\ f(x)=1}} \int_\Omega |f(\xi)|^2\, d\mu(\xi), \quad x \in \Omega, \tag{3.6}$$

where it is agreed that $\lambda(X_N, \mu; x) = \infty$ if $f(x) = 0$ for all $f \in X_N$. It is easily seen from (1.6) that

$$\int_\Omega \frac{1}{\lambda(X_N, \mu; x)} d\mu(x) = N.$$

Using this notation, we now derive a weighted discretization result for more general spaces X_N without imposing the $\mathrm{NI}_{2,\infty}(\sqrt{KN})$ assumption.

Corollary 3.8. *Let X_N be an N-dimensional subspace of $L_2(\Omega,\mu)$ with the associated Christoffel function $\lambda_N(\cdot) = \lambda(X_N, \mu; \cdot)$ as defined in (3.6). Let $\xi^1, \dots, \xi^m$ be a sequence of IID random points drawn from the distribution $\frac{1}{N\lambda_N(x)}\,\mu(dx)$. If $m \geqslant 3AN\epsilon^{-2}\log N$ for some constants $A > 1$ and $\epsilon \in (0,1)$, then the inequalities*

$$(1-\epsilon)\|f\|^2_{L_2(\Omega,\mu)} \leqslant \frac{N}{m}\sum_{j=1}^{m} \lambda_N(\xi^j)\left|f\left(\xi^j\right)\right|^2 \leqslant (1+\epsilon)\|f\|^2_{L_2(\Omega,\mu)}$$

hold simultaneously for all $f \in X_N$ with probability at least $1 - 2N^{-A+1}$.

Proof. Let $\{u_1, \dots, u_N\}$ be an orthonormal basis of X_N. By (1.6), we have

$$\frac{1}{\lambda_N(x)} = |F(x)|^2 \quad \text{with} \quad F(x) := \big(\sum_{k=1}^{N} |u_k(x)|^2 \big)^{1/2}.$$

Consider the probability measure $d\nu := N^{-1}F^2 d\mu$ on Ω. Define a mapping $U : L_2(\Omega, \mu) \to L_2(\Omega, \nu)$ by $Uf(x) = N\, f(x)/F(x)$ if $F(x) > 0$, and $Uf(x) = 0$ otherwise. Since $F(x) = 0$ implies $f(x) = 0$ for all $f \in X_N$, it follows that

$$\|Uf\|_{L_2(\Omega,\nu)} = \|f\|_{L_2(\Omega,\mu)}, \quad \forall f \in X_N.$$

Moreover, it is easily seen from (3.6) that

$$\sup_{x\in\Omega} |g(x)| \leqslant \sqrt{N}\|g\|_{L_2(\Omega,\nu)}, \quad \forall g \in UX_N.$$

Corollary 3.8 then follows by applying Theorem 3.2 to the space $UX_N \subset L_2(\Omega, \nu)$ with $K = 1$. □

3.3 The minimum number of points required for L_2-discretization

Given two constants $C_2 \geqslant C_1 > 0$, we denote by $m(X_N; 2; C_1, C_2)$ the smallest positive integer m for which there exist m points $\xi^1, \dots, \xi^m \in \Omega$ such that

$$C_1\|f\|^2_{L_2(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^2 \leqslant C_2\|f\|^2_{L_2(\Omega,\mu)}, \quad \forall f \in X_N.$$

Clearly, $m(X_N; 2; C_1, C_2) \geqslant N$. Regarding the upper bound, Theorem 3.2 (originally due to Rudelson [55]) establishes that if $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$, then

$$m(X_N; 2; 1/2, 3/2) \leqslant CKN \log N.$$

The following remarkable theorem asserts that the extra factor $\log N$ in this bound can be removed. The proof relies on the breakthrough theorem of A. Marcus, D. Spielman, and N. Srivastava [50] combined with an iterative procedure initially proposed in [7, 49].

Theorem 3.9 ([37, 46])**.** *Let X_N be an N-dimensional space of bounded functions on Ω such that $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ for some constant $K \geqslant 1$. Then for any $\epsilon \in (0, 1)$, there exist m points $\xi^1, \dots, \xi^m \in \Omega$ with $m \leqslant C\epsilon^{-2}KN$ for some absolute constant $C > 0$, such that*

$$(1-\epsilon)\|f\|^2_{L_2(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^2 \leqslant (1+\epsilon)K\|f\|^2_{L_2(\Omega,\mu)}, \quad \forall f \in X_N. \tag{3.7}$$

We point out that Theorem 3.9 was established in [46] with general positive constants C_1 and C_2K instead of the constants $1 - \epsilon$ and $(1 + \epsilon)K$ in (3.7). The current version of Theorem 3.9 is due to [37], where it was observed that inequality (3.7) can be obtained by refining the iteration technique introduced in [46].

By combining Theorem 3.9 with the change of density argument from the proof of Corollary 3.8, we obtain an immediate bound on the required number of evaluation points.

Corollary 3.10 ([37]). *Given any N-dimensional space X_N of bounded functions on Ω, and any $\epsilon \in (0,1)$, there exist points $\xi^1, \dots, \xi^m \in \Omega$ and weights $\lambda_1, \dots, \lambda_m > 0$ with $m \leqslant C\epsilon^{-2}N$ for some absolute constant $C > 0$, such that*

$$(1-\epsilon)\|f\|^2_{L_2(\Omega,\mu)} \leqslant \sum_{j=1}^m \lambda_j |f(\xi^j)|^2 \leqslant (1+\epsilon)\|f\|^2_{L_2(\Omega,\mu)}, \quad \forall f \in X_N.$$

Example 3.11. Let $Q \subset \mathbb{Z}^d$ be a finite set of frequencies, and consider the space

$$\mathcal{T}(Q) := \operatorname{span}\{e^{\mathrm{i}\langle k,\cdot\rangle} : k \in Q\}$$

of all trigonometric polynomials with frequencies from Q on the cube $\Omega = [0, 2\pi)^d$, equipped with the normalized Lebesgue measure μ. Since the moduli of functions in the orthonormal basis $\{e^{\mathrm{i}\langle k,\cdot\rangle} : k \in Q\}$ of $\mathcal{T}(Q)$ are uniformly bounded by 1, the Cauchy-Schwarz inequality implies that $\mathcal{T}(Q) \in \mathrm{NI}_{2,\infty}(\sqrt{N})$ with $N := \dim \mathcal{T}(Q) = |Q|$. Therefore, by Theorem 3.9, we can find a set of $m \leqslant C\epsilon^{-2}N$ points $\xi^1, \dots, \xi^m \in \Omega$ such that

$$(1-\epsilon)\|f\|^2_{L_2(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^2 \leqslant (1+\epsilon)\|f\|^2_{L_2(\Omega,\mu)}, \quad \forall f \in \mathcal{T}(Q).$$

We conclude this section with the following tight estimate on weighted Marcinkiewicz discretization inequality of L_2 norms, which follows directly from a result of G. Schechtman [57].

Theorem 3.12 ([57]). *If X_N is an N-dimensional subspace of $L_2(\Omega,\mu)$, then for any $b \in (1,2]$, there exist a set of $m \leqslant bN$ points $\xi^1, \dots, \xi^m \in \Omega$ and a set of nonnegative weights $\lambda_1, \dots, \lambda_m$ such that*

$$\|f\|_{L_2(\Omega,\mu)} \leqslant \left(\sum_{j=1}^m \lambda_j \left|f\left(\xi^j\right)\right|^2\right)^{1/2} \leqslant \frac{C}{b-1}\|f\|_{L_2(\Omega,\mu)}, \quad \forall f \in X_N,$$

where $C > 1$ is an absolute constant.

Note that in Theorem 3.12, the parameter $b > 1$ can be chosen to be arbitrarily close to 1.

4 Marcinkiewicz discretization theorems

This section presents general results and methods concerning the discretization of L_p norms for $p \neq 2$.

4.1 A conditional theorem

Let $\mathcal{W}$ be a compact subset of the space of bounded functions on Ω. Let $\xi^1, \xi^2, \dots$ be a sequence of independent random points in Ω. Consider the following random sampling discretization problem: for $\epsilon \in (0,1)$, $1 \leqslant p < \infty$ and $f \in \mathcal{W}$,

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}. \tag{4.1}$$

Without loss of generality, we may assume that

$$\mathcal{W} \subset B_{L_p} := \big\{ f \in L_p(\Omega, \mu) : \|f\|_{L_p(\Omega,\mu)} \leqslant 1 \big\}$$

since otherwise we may replace $\mathcal{W}$ with the set $\big\{ f/\|f\|_{L_p(\Omega,\mu)} \colon f \in \mathcal{W} \big\}$. Our goal is to estimate the probability that (4.1) holds for all $f \in \mathcal{W}$ in terms of the number m of required random points. Such an estimate can be established in terms of an integral of the ϵ-entropy of the set $\mathcal{W}$ in the uniform metric. We begin by recalling standard definitions and known estimates concerning entropy (see [9, 26, 59] and [62, Chapter 7]).

Entropy. Consider a normed linear space $X := (X, \|\cdot\|_X)$, and let

$$B_X := \big\{ x \in X \colon \|x\|_X \leqslant 1 \big\}$$

denote the closed unit ball in X. For any $\epsilon > 0$, the covering number $N_\epsilon(A, X)$ of a compact subset $A \subset X$ is defined as the smallest positive integer n for which there exist elements $x^1, \dots, x^n \in A$ such that

$$A \subset \bigcup_{j=1}^{n} \big(x^j + \epsilon B_X \big).$$

The ϵ-entropy $\mathcal{H}_\epsilon(A; X)$ of a compact set A in X is defined by

$$\mathcal{H}_\epsilon(A, X) := \log_2 N_\epsilon(A, X),$$

while the entropy numbers $e_k(A, X)$ of the set A in X are defined by

$$e_k(A, X) := \inf\{\epsilon > 0 \colon \mathcal{H}_\epsilon(A, X) \leqslant k\}, \qquad k = 0, 1, 2, \dots.$$

Note that in our definition we require $x^j \in A$, which differs from the standard definition of $N_\epsilon(A, X)$ and $e_k(A, X)$, where this restriction is not imposed. However, it is well known (see [62], p. 208) that these characteristics may differ by at most a factor of 2. To emphasize the dependence on the seminorm $\|\cdot\|_X$, we often write $e_k(A, \|\cdot\|_X) = e_k(A, X)$ and $\mathcal{H}_\epsilon(A, \|\cdot\|_X) = \mathcal{H}_\epsilon(A, X)$.

There is a simple and useful bound for the entropy of the unit ball in the finite-dimensional case. If $X = \mathbb{R}^m$, then for any $0 < \epsilon \leqslant 1$,

$$m \log_2 \frac{1}{\epsilon} \leqslant \mathcal{H}_\epsilon(B_X, \|\cdot\|_X) \leqslant m \log_2 \Big(1 + \frac{2}{\epsilon} \Big). \tag{4.2}$$

This also implies that

$$2^{-1} \cdot 2^{-k/m} \leqslant e_k(B_X, \|\cdot\|_X) \leqslant 6 \cdot 2^{-k/m}, \quad k \in \mathbb{N}.$$

Sharp entropy estimates are generally difficult to establish in a general high-dimensional setting. Here we only formulate a useful bound for entropy numbers, which will be used later.

Theorem 4.1 ([36, Lemma 4.10, Corollary 4.2])**.** *Let $\Omega_m = \{\xi^1, \dots, \xi^m\} \subset \Omega$ be a finite subset of points in Ω. Let X_N be an N-dimensional space of bounded functions on Ω satisfying*

$$\|f\|_{L_\infty(\Omega)} \leqslant \sqrt{M} \|f\|_{L_2(\Omega,\mu)}, \quad \forall f \in X_N, \tag{4.3}$$

where $M \geqslant 2$ is a constant. Then for any $1 \leqslant p \leqslant 2$,

$$e_k(X_N^p, \|\cdot\|_{L_\infty(\Omega_m)}) \leqslant C_p (\log m)^{\frac{1}{2}} (\log M)^{\frac{1}{p} - \frac{1}{2}} \Big(\frac{M}{k} \Big)^{\frac{1}{p}}, \quad k = 0, 1, 2, \dots \tag{4.4}$$

where $X_N^p := \{ f \in X_N \colon \|f\|_{L_p(\Omega,\mu)} \leqslant 1 \}$.

Remark 4.2. For $p \in (1,2)$, Theorem 4.1 was established in [36, Lemma 4.10] via a slight modification of the proof of Proposition 16.8.6 in [59]. The case $p = 2$ is an immediate consequence of [36, Corollary 4.2]. Finally, for $p = 1$, the conclusion follows from [16, (3.8) and (3.10)] and the following estimate on entropy numbers (see, for instance, [62, p. 323, (7.1.6)]):

$$e_{2k}(X_N^1, \|\cdot\|_{L_\infty(\Omega_m)}) \leqslant C e_k(X_N^1, \|\cdot\|_{L_2(\Omega,\mu)}) e_k(X_N^2, \|\cdot\|_{L_\infty(\Omega_m)}).$$

Remark 4.3. While the dimension N does not appear explicitly in the estimate (4.4), the constant M in the Nikolski inequality 4.3 typically depends on N, particularly when $m \gg N$. Furthermore, the use of the $L_\infty(\Omega_m)$ norm embeds the original space X_N into a subspace of $\mathbb{R}^m$.

Remark 4.4. By definition, it is easily seen that the estimates of entropy numbers

$$e_k(A, X) \leqslant Rk^{-\alpha}, \quad k = 1, 2, \dots$$

imply

$$\mathcal{H}_\epsilon(A, X) \leqslant c(R/\epsilon)^{1/\alpha}, \quad \epsilon > 0,$$

where $R > 0$ is a constant independent of $k \in \mathbb{N}$, and $c > 0$ is an absolute constant. As a result, Theorem 4.1 implies that under the condition (4.3), for any $1 \leqslant p \leqslant 2$, we have

$$\mathcal{H}_\epsilon(X_N^p, \|\cdot\|_{L_\infty(\Omega_m)}) \leqslant C_p (\log m)^{\frac{p}{2}} (\log M)^{1-\frac{p}{2}} M \epsilon^{-p}, \quad \epsilon > 0.$$

We are now ready to formulate the conditional discretization theorem for the Marcinkiewicz-type discretization inequality of the form (4.1). The following theorem extends results in [17, Theorem 1.3], [20, Theorem 5.1] and [61].

Theorem 4.5 ([21, Theorem 6.1])**.** *Let $(\Omega, \mu, \mathcal{F})$ be a probability space. Let $\mathcal{W}$ be a compact set of bounded functions on Ω satisfying the following conditions for some $1 \leqslant p < \infty$:*

(i) $\sup\limits_{f\in\mathcal{W}} \sup\limits_{x\in\Omega} |f(x)| \leqslant R^{1/p}$ *for some constant* $R \geqslant 1$*;*

(ii) $\{\lambda f : \lambda > 0, f \in \mathcal{W},\ \|f\|_{L_p(\Omega,\mu)} \leqslant 1\} = \mathcal{W}$.

Let $\xi^1, \dots, \xi^m$ be a sequence of independent random points on Ω satisfying

$$\sum_{k=1}^m \mathbb{P}[\xi^k \in E] = m \cdot \mu(E), \quad \forall E \in \mathcal{F}. \tag{4.5}$$

Then there exist positive constants $C_p \geqslant 1$ and $c_p \in (0,1)$ depending only on p such that for any $\epsilon \in (0, 1/2)$ and any integer

$$m \geqslant C_p \epsilon^{-5} \Big(\log \frac{R}{\epsilon}\Big) \int_{c_p\epsilon}^R \mathcal{H}_{\epsilon t^{1/p}}(\mathcal{W}; L_\infty) dt, \tag{4.6}$$

the inequality

$$(1-\epsilon)\|f\|_{L_p(\Omega,\mu)}^p \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|_{L_p(\Omega,\mu)}^p$$

holds simultaneously for all $f \in \mathcal{W}$ with probability at least

$$1 - \exp\Big(-\frac{c_p m \epsilon^4}{R(\log \frac{R}{\epsilon})^2}\Big).$$

Remark 4.6. Theorem 4.5 was proved in [21, Theorem 6.1], where (4.6) is replaced by the following slightly different condition:

$$m \geqslant C_p \epsilon^{-5} \left(\int_{c_p \epsilon^{1/p}}^{R^{1/p}} u^{\frac{p}{2}-1} \Big(\int_u^{R^{1/p}} \frac{\mathcal{H}_{\epsilon t}(\mathcal{W}; L_\infty)}{t}\, dt \Big)^{\frac{1}{2}} du \right)^2. \tag{4.7}$$

Note that, performing the change of variables $v = u^p$ and $s = t^p$ yields

$$\text{RHS of (4.7)} = C'_p \epsilon^{-5} \left(\int_{c'_p \epsilon}^{R} v^{-\frac{1}{2}} \Big(\int_v^R \frac{\mathcal{H}_{\epsilon s^{1/p}}(\mathcal{W}; L_\infty)}{s}\, ds \Big)^{\frac{1}{2}} dv \right)^2,$$

where $c'_p \in (0,1)$. Using the Cauchy–Bunyakovsky–Schwarz inequality and the Fubini theorem, we then obtain

$$\text{RHS of (4.7)} \leqslant C_p \epsilon^{-5} \Big(\log \frac{R}{\epsilon} \Big) \int_{c'_p \epsilon}^{R} \mathcal{H}_{\epsilon t^{1/p}}(\mathcal{W}; L_\infty) dt.$$

Remark 4.7. Either of the following two conditions implies the condition (4.5) in Theorem 4.5:

(i) $\xi^1, \dots, \xi^m$ are identically distributed according to μ;

(ii) there exists a partition $\{\Lambda_1, \dots, \Lambda_m\}$ of Ω such that $\mu(\Lambda_j) = \frac{1}{m}$ and $\xi^j \in \Lambda_j$ is distributed according to $m \cdot \mu\Big|_{\Lambda_j}$ for each $1 \leqslant j \leqslant m$.

Theorem 4.5 combined with Theorem 4.1 implies Corollary 4.8 below.

Corollary 4.8. *Let $\{V_j\}_{j=1}^L$ be a sequence of linear subspaces of $\mathbb{R}^n$ which satisfies the following condition for some constant $R \geqslant 1$:*

$$\|f\|_{L_\infty^n} \leqslant \sqrt{R} \|f\|_{L_2^n}, \quad \forall f \in \bigcup_{j=1}^L V_j. \tag{4.8}$$

Let $\xi^1, \dots, \xi^m$ be a sequence of independent random variables taking values in $\{1, 2, \dots, n\}$ and satisfying

$$\frac{1}{m} \sum_{j=1}^m \mathbb{P}[\xi^j \in E] = \frac{|E|}{n}, \quad \forall E \subset \{1, 2, \dots, n\}.$$

Let $\epsilon \in (0, \frac{1}{2})$. Then the following statements hold.

(i) *If $1 \leqslant p \leqslant 2$, then there exist constants $C_p \geqslant 1$ and $c_p > 0$ such that for any integer*

$$m \geqslant C_p \epsilon^{-5} R \Big(\log \frac{R}{\epsilon} \Big) \Big[\epsilon^{-p} \Big(\log \frac{R}{\epsilon} \Big) (\log R)^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}} + \log L \Big],$$

the Marcinkiewicz-type discretization inequality

$$(1-\epsilon) \|f\|_{L_p^n}^p \leqslant \frac{1}{m} \sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon) \|f\|_{L_p^n}^p \tag{4.9}$$

holds simultaneously for all $f \in \bigcup_{j=1}^L V_j$ with probability at least

$$1 - \exp\Big(- \frac{c_p m \epsilon^4}{R (\log \frac{R}{\epsilon})^2} \Big).$$

(ii) *If $2 < p < \infty$, then there exist constants $C_p \geqslant 1$ and $c_p > 0$ such that for any integer*

$$m \geqslant C_p \epsilon^{-5} R^{p/2} \left(\log \frac{R}{\epsilon}\right)\Big[\epsilon^{-2}\log n + \log L\Big],$$

the Marcinkiewicz-type discretization inequality (4.9) *holds simultaneously for all $f \in \bigcup_{j=1}^{L} V_j$ with probability at least*

$$1 - \exp\Big(-\frac{c_p m \epsilon^4}{R^{p/2}(\log \frac{R}{\epsilon})^2}\Big).$$

Proof. Let $\Omega_n := \{1, 2, \dots, n\}$ and let μ_n denote the uniform distribution on Ω_n. Then the sequence of random variables ξ^j satisfies the condition (4.5) with $\Omega = \Omega_n$ and $\mu = \mu_n$. Let

$$\mathcal{W}_p := \Big\{ f \in \bigcup_{j=1}^{L} V_j : \ \|f\|_{L_p^n} \leqslant 1\Big\}.$$

(i) Let $1 \leqslant p \leqslant 2$. Using Theorem 4.1 with $M = R$, $m = n$, and $\Omega_m = \Omega_n$, we obtain

$$e_k(V_j^p, \|\cdot\|_{L_\infty^n}) \leqslant C_p (\log n)^{\frac{1}{2}} \ (\log R)^{\frac{1}{p}-\frac{1}{2}} \Big(\frac{R}{k}\Big)^{\frac{1}{p}}, \quad k \in \mathbb{N}, \quad j = 1, 2, \dots, L,$$

where $V_j^p := \{f \in V_j : \ \ \|f\|_{L_p^n} \leqslant 1\}$. By Remark 4.4, this implies

$$\mathcal{H}_t(V_j^p; \|\cdot\|_{L_\infty^n}) \leqslant C_p t^{-p} R (\log R)^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}}, \quad \forall t > 0.$$

Since $\mathcal{W}_p = \bigcup_{j=1}^{L} V_j^p$, it follows that for any $t > 0$,

$$\begin{aligned}\mathcal{H}_t(\mathcal{W}_p; \|\cdot\|_{L_\infty^n}) &\leqslant \log_2 L + \max_{1\leqslant j\leqslant L} \mathcal{H}_t(V_j^p; \|\cdot\|_{L_\infty^n}) \\ &\leqslant C_p t^{-p} R (\log R)^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}} + 2\log L. \end{aligned} \tag{4.10}$$

Thus,

$$\begin{aligned}\int_{c_p\epsilon}^{R} \mathcal{H}_{\epsilon t^{1/p}}(\mathcal{W}_p; \|\cdot\|_{L_\infty^n}) dt &\leqslant C_p \epsilon^{-p} R (\log R)^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}} \int_{c_p\epsilon}^{R} t^{-1}\, dt + 2R\log L \\ &\leqslant C_p \epsilon^{-p} \Big(\log \frac{R}{\epsilon}\Big) R (\log R)^{1-\frac{p}{2}} (\log n)^{\frac{p}{2}} + 2R\log L.\end{aligned}$$

Finally, by (1.8), the assumption (4.8) implies that $\|f\|_{L_\infty^n} \leqslant R^{1/p}\|f\|_{L_p^n}$ for all $f \in \bigcup_{j=1}^{L} V_j$. Consequently, Statement (i) follows directly by applying Theorem 4.5 to $\mathcal{W} = \mathcal{W}_p$.

(ii) Now assume that $p > 2$. Then for all $f \in \bigcup_{j=1}^{L} V_j$, we have

$$\|f\|_{L_\infty^n} \leqslant \sqrt{R}\|f\|_{L_2^n} \leqslant R_1^{1/p}\|f\|_{L_p^n} \ \text{ with } \ R_1 = R^{p/2}.$$

Since $\mathcal{W}_p \subset \mathcal{W}_2$, we obtain from (4.10) that for any $t > 0$,

$$\mathcal{H}_t(\mathcal{W}_p; \|\cdot\|_{L_\infty^n}) \leqslant \mathcal{H}_t(\mathcal{W}_2; \|\cdot\|_{L_\infty^n}) \leqslant C t^{-2} R(\log n) + 2\log L,$$

implying that

$$\begin{aligned}\int_{c_p\epsilon}^{R_1} \mathcal{H}_{\epsilon t^{1/p}}(\mathcal{W}_p; \|\cdot\|_{L_\infty^n}) dt &\leqslant C\epsilon^{-2} R\log n \int_{c_p\epsilon}^{R^{p/2}} t^{-\frac{2}{p}}\, dt + 2R^{p/2}\log L \\ &\leqslant C_p R^{p/2}\Big[\epsilon^{-2}\log n + \log L\Big].\end{aligned}$$

Statement (ii) then follows by applying Theorem 4.5 to $\mathcal{W} = \mathcal{W}_p$. □

4.2 Preliminary discretizations

Theorem 4.5 is often applied together with the entropy estimates stated in Theorem 4.1, as was done in the proof of Corollary 4.8. This normally requires a preliminary step of discretization involving more random points. The following result is useful.

Proposition 4.9. *Let $1 \leqslant p < \infty$ be a fixed number. Let $\{V_j\}_{j=1}^L$ be a sequence of linear subspaces of $L_\infty(\Omega)$ such that $\dim V_j \leqslant s$ for all $1 \leqslant j \leqslant L$, and*

$$\|f\|_{L_\infty(\Omega)} \leqslant R^{1/p}\|f\|_{L_p(\Omega,\mu)}, \quad \forall f \in \bigcup_{j=1}^L V_j, \tag{4.11}$$

for some constant $R \geqslant 1$. Assume that $0 < \epsilon \leqslant \frac{1}{8}$, that m is an integer satisfying

$$m \geqslant 16R\epsilon^{-2}\Big[\log L + s\log\frac{3}{\epsilon}\Big], \tag{4.12}$$

and that $\{\xi^j\}_{j=1}^m$ is a sequence of IID random points drawn from a probability distribution μ on Ω. Then the Marcinkiewicz-type discretization inequality

$$(1-5\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+5\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \tag{4.13}$$

holds simultaneously for all $f \in \bigcup_{j=1}^L V_j$ with probability $\geqslant 1 - 2\exp\Big(-\frac{m\epsilon^2}{16R}\Big)$.

We point out that in the above proposition, the upper bound $\frac{1}{8}$ in the assumption $0 < \epsilon \leqslant \frac{1}{8}$ is chosen purely for convenience, any value in $(0, \frac{1}{5})$ works by appropriately modifying the constant 16 in (4.12).

For the proof of Proposition 4.9, we need the following two lemmas. The first lemma appears in [7, Lemma 2.1], but also follows directly from the classical Bernstein inequality (see [65, Theorem 4.22]).

Lemma 4.10 ([7, Lemma 2.1])**.** *Let $\{\eta_j\}_{j=1}^m$ be an independent sequence of random variables such that $\mathbb{E}\eta_j = 0$, $\mathbb{E}|\eta_j| \leqslant M_1$ and $|\eta_j| \leqslant M_\infty$ almost surely for all $1 \leqslant j \leqslant m$ and some constants $M_1, M_\infty > 0$. Then for any $0 < \lambda < M_1$,*

$$\mathbb{P}\Big[|\sum_{j=1}^m \eta_j| \geqslant m\lambda\Big] \leqslant 2e^{-\frac{m\lambda^2}{4M_1M_\infty}}.$$

A slight variant of the second lemma also appears in [7, Lemma 2.5].

Lemma 4.11 ([7, Lemma 2.5])**.** *Let $T : X \to Y$ be a bounded linear operator between two normed linear spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$. Let $\epsilon \in (0, \frac{1}{2})$. Assume that there exists an ϵ-net $\mathcal{A} \subset B_X$ of B_X such that for some constants $\alpha, \beta > 0$,*

$$\alpha\|x\|_X \leqslant \|Tx\|_Y \leqslant \beta\|x\|_X, \quad \forall x \in \mathcal{A}.$$

Then, with $\gamma = \dfrac{3\beta}{\alpha}$, we have

$$\alpha(1-\gamma\epsilon)\|z\|_X \leqslant \|Tz\|_Y \leqslant \beta(1+2\epsilon)\|z\|_X, \quad \forall z \in X.$$

For completeness, we provide a proof of Lemma 4.11 below.

Proof. For $z \in B_X$, we denote by $\varphi(z)$ an element in $\mathcal{A}$ such that $\|z - \varphi(z)\|_X \leqslant \epsilon$. Then for any $z \in B_X$, we have

$$\|Tz\|_Y \leqslant \|T\|\epsilon + \|T\varphi(z)\|_Y \leqslant \|T\|\epsilon + \beta\|\varphi(z)\|_X \leqslant \|T\|\epsilon + \beta,$$

which, taking supremum over all $z \in B_X$, implies

$$\|T\| \leqslant \frac{\beta}{1-\epsilon} \leqslant \beta(1+2\epsilon).$$

To prove the left-hand-side estimate, we note that for any $z \in X$ with $\|z\|_X = 1$, we have

$$\begin{aligned}\|Tz\|_Y &\geqslant \|T\varphi(z)\|_Y - \|T\|\epsilon \geqslant \alpha\|\varphi(z)\|_X - \|T\|\epsilon \\ &\geqslant \alpha(1-\epsilon) - 2\beta\epsilon \geqslant \alpha - 3\beta\epsilon.\end{aligned}$$

This proves the lemma. □

Proof of Proposition 4.9 . By inequality (4.2), for each $1 \leqslant j \leqslant L$, there exists an ϵ-net $\mathcal{A}_j \subset V_j^p$ of $V_j^p := V_j \cap B_{L_p}$ in the space L_p such that $|\mathcal{A}_j| \leqslant \left(1 + \frac{2}{\epsilon}\right)^s$. Let $\mathcal{A} := \bigcup_{j=1}^L \mathcal{A}_j$. Then

$$\log|\mathcal{A}| \leqslant \log\Big[L\Big(1+\frac{2}{\epsilon}\Big)^s\Big] \leqslant \log L + s\log\frac{3}{\epsilon}. \tag{4.14}$$

Furthermore, (4.11) implies that

$$\frac{\|f\|_\infty^p}{\|f\|_p^p} = \left\|\frac{|f|^p}{\|f\|_p^p}\right\|_\infty \leqslant R \quad \text{for all } f \in \mathcal{A},$$

where $\|f\|_\infty = \sup_{x\in\Omega}|f(x)|$ and $\|f\|_p = \|f\|_{L_p(\Omega,\mu)}$. Thus, using (4.14), and Lemma 4.10 with $M_\infty = R$, $M_1 = 2$ and

$$\eta_j = \frac{|f(\xi^j)|^p}{\|f\|_p^p} - 1, \quad j = 1, 2, \dots, m,$$

we conclude that the inequalities

$$\Big|\frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p - \int_\Omega |f|^p\,d\mu\Big| \leqslant \epsilon\|f\|_{L_p(\Omega,\mu)}^p, \quad \forall f \in \mathcal{A} \tag{4.15}$$

hold with probability

$$\begin{aligned}&\geqslant 1 - 2|\mathcal{A}|\exp\Big(-\frac{m\epsilon^2}{8R}\Big) = 1 - 2\exp\Big(\log|\mathcal{A}| - \frac{m\epsilon^2}{8R}\Big) \\ &\geqslant 1 - 2\exp\Big(\log L + s\log\frac{3}{\epsilon} - \frac{m\epsilon^2}{8R}\Big).\end{aligned}$$

However, the condition (4.12) implies

$$\log L + s\log\frac{3}{\epsilon} \leqslant \frac{m\epsilon^2}{16R}.$$

It follows that (4.15) holds for all $f \in \mathcal{A}$ with probability

$$\geqslant 1 - 2\exp\Big(-\frac{m\epsilon^2}{16R}\Big).$$

To complete the proof, we just need to note that by Lemma 4.11, (4.15) implies (4.13). □

4.3 Sampling discretization of L_p norms

In this subsection, we show how to use the preliminary discretization from the last subsection, together with Corollary 4.8, to deduce a Marcinkiewicz-type discretization theorem.

Theorem 4.12. *Let X_N be an N-dimensional space of bounded functions on Ω, and assume that $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ for some constant $K \geqslant 2$. Let $\{\xi^j\}_{j=1}^{\infty}$ be a sequence of IID random points drawn from the distribution μ. Let $\epsilon \in (0, 1/2)$.*

(i) *If $1 \leqslant p \leqslant 2$, then there exists a constant $C_p(\epsilon) \geqslant 1$ such that for any given parameter $A \geqslant 1$, and any integer*

$$m \geqslant C_p(\epsilon) A \cdot KN \cdot \log^3(2KN), \tag{4.16}$$

the Marcinkiewicz-type discretization inequality

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}$$

holds simultaneously for all $f \in X_N$ with probability $\geqslant 1 - (KN)^{-A}$.

(ii) *If $2 < p < \infty$, then there exists a constant $C_p(\epsilon) \geqslant 1$ such that for any given parameter $A \geqslant 1$, and any integer*

$$m \geqslant C_p(\epsilon) A \cdot (KN)^{p/2} \cdot \log^2(2KN),$$

the Marcinkiewicz-type discretization inequality

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}$$

holds simultaneously for all $f \in X_N$ with probability $\geqslant 1 - e^{-A}$.

Proof. We start with the proof of (i), when $1 \leqslant p \leqslant 2$. The assumption $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ implies (see (1.9)) that

$$\|f\|_{L_\infty(\Omega)} \leqslant (KN)^{1/p}\|f\|_{L_p(\Omega,\mu)}, \quad f \in X_N.$$

Thus, without loss of generality, we may assume that $m \leqslant CKN^2\epsilon^{-2}\log\frac{1}{\epsilon}$, where $C > 0$ is a sufficiently large fixed constant, since otherwise Theorem 4.12 follows directly from Proposition 4.9 with $L = 1$, $s = N$, and $R = KN$. We will use Part (i) of Corollary 4.8. To this end, we need a step of preliminary discretization involving more points.

Let ℓ_N denote the smallest integer larger than or equal to $CKN^2\epsilon^{-2}\log\frac{1}{\epsilon}$. Let $m_1 := m\ell_N$, and let $\Omega_{m_1} := \{x_1, \dots, x_{m_1}\} \subset \Omega$ be a set of IID random points drawn from the distribution μ. Note that

$$C_1(C,\epsilon)\log(KN) \leqslant \log m_1 \leqslant C_2(C,\varepsilon)\log(KN)$$

for some constants $C_2(C,\epsilon) \geqslant C_1(C,\epsilon) > 0$. Furthermore, by Proposition 4.9, the following inequalities hold simultaneously with probability $\geqslant 1 - 4\exp\left(-\frac{cm_1\epsilon^2}{KN}\right) \geqslant 1 - 4e^{-mN}$, for a sufficiently large constant $C > 0$,

$$\left|\|f\|^p_{L_p(\Omega,\mu)} - \|f\|^p_{L_p(\Omega_{m_1})}\right| \leqslant \frac{\epsilon}{10}\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in X_N, \tag{4.17}$$

$$\left|\|f\|^2_{L_2(\Omega,\mu)} - \|f\|^2_{L_2(\Omega_{m_1})}\right| \leqslant \frac{\epsilon}{10}\|f\|^2_{L_2(\Omega,\mu)}, \quad \forall f \in X_N. \tag{4.18}$$

Let $\{\Lambda_1, \dots, \Lambda_m\}$ be a partition of the set $\{1, \dots, m_1\}$ such that $|\Lambda_j| = \ell_N$ for $1 \leqslant j \leqslant m$. Let $n_1, \dots, n_m$ be an independent sequence of random variables that is independent of $\{x_1, \dots, x_{m_1}\}$ such that each n_j is uniformly distributed in Λ_j. Define $\xi^j = x_{n_j}, \quad j = 1, 2, \dots, m$. Then $\xi^1, \dots, \xi^m \in \Omega$ are IID random points with common distribution μ.

Next, we fix a set Ω_{m_1} of random points $x_1, \dots, x_{m_1}$ for which both (4.17) and (4.18) hold. Instead of dealing with functions on the entire set Ω, we consider their restrictions on the set Ω_{m_1} equipped with the uniform distribution. From the assumption $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ and by (4.18), we have

$$\|f\|_{L_\infty(\Omega_{m_1})} \leqslant \sqrt{2KN}\|f\|_{L_2(\Omega_{m_1})}, \quad \forall f \in X_N.$$

Thus, applying Part (i) of Corollary 4.8 with $L = 1$, $n = m_1$ and $R = 2KN$, we conclude that the inequalities

$$\Big|\frac{1}{m}\sum_{j=1}^{m} |f(x_{n_j})|^p - \|f\|^p_{L_p(\Omega_{m_1})}\Big| \leqslant \frac{\epsilon}{10}\|f\|^p_{L_p(\Omega_{m_1})}, \quad \forall f \in X_N \tag{4.19}$$

hold with probability

$$\geqslant 1 - \exp\Big(-\frac{c_p(\epsilon)m}{KN\log^2(KN)}\Big)$$

provided that

$$m \geqslant C_p(\epsilon)KN\log^3(KN).$$

Now, we recall that $\xi^j = x_{n_j}$. Clearly, (4.19) combined with (4.17) yields that

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in X_N. \tag{4.20}$$

To conclude the proof, we estimate the probability of the event E that (4.20) holds. Let E_1 denote the event that both (4.17) and (4.18) hold. Note that χ_{E_1} is a function of $(x_1, x_2, \dots, x_{m_1})$, and $\mathbb{P}(E_1) \geqslant 1 - 4e^{-mN}$. Moreover, from the above proof, we have

$$\chi_{E_1} \cdot \mathbb{E}[\chi_E | x_1, \dots, x_{m_1}] \geqslant \chi_{E_1} \cdot \left(1 - \exp\Big(-\frac{c_p(\epsilon)m}{KN\log^2(KN)}\Big)\right),$$

implying that

$$\begin{aligned}\mathbb{E}\Big[\chi_{E_1} \cdot \mathbb{E}[\chi_E | x_1, \dots, x_{m_1}]\Big] &\geqslant \mathbb{P}(E_1) \cdot \left(1 - \exp\Big(-\frac{c_p(\epsilon)m}{KN\log^2(KN)}\Big)\right) \\ &\geqslant 1 - 5\exp\Big(-\frac{c_p(\epsilon)m}{KN\log^2(KN)}\Big).\end{aligned}$$

Thus, if m is an integer satisfying (4.16) for some real $A \geqslant 1$, then

$$\begin{aligned}\mathbb{P}(E) \geqslant \mathbb{E}[\chi_{E\cap E_1}] &= \mathbb{E}\Big[\chi_{E_1} \cdot \mathbb{E}[\chi_E | x_1, \dots, x_{m_1}]\Big] \\ &\geqslant 1 - 5\exp\Big(-C_p(\epsilon)c_p(\epsilon)A\log(KN)\Big) \geqslant 1 - (KN)^{-A}\end{aligned}$$

provided that the constant $C_p(\varepsilon) > 0$ is sufficiently large. This completes the proof of (i).

The proof of Part (ii) is almost identical to that of Part (i), with all instances of $KN^2\varepsilon^{-2}\log\frac{1}{\varepsilon}$ replaced by $K^{p/2}N^{1+p/2}\varepsilon^{-2}\log\frac{1}{\varepsilon}$. The only difference is that, after the preliminary discretization step, we apply Part (ii) of Corollary 4.8 instead of Part (i), together with the elementary inequality

$$\|f\|_{L_2(\Omega,\mu)} \leqslant \|f\|_{L_p(\Omega,\mu)}, \qquad p > 2.$$

Since no new ideas are involved, we omit the details. □

4.4 Sampling discretization with weights

The constant K in the $(2,\infty)$-Nikol'skii condition $\mathrm{NI}_{2,\infty}(\sqrt{KN})$ can grow substantially as the dimension of X_N increases. For instance, if X_N is the space of all algebraic polynomials of degree less than N on the interval $[-1,1]$ and μ is the probability measure on $[-1,1]$ given by

$$d\mu(x) = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda+\frac{1}{2})\sqrt{\pi}}(1-x^2)^{\lambda-\frac{1}{2}}\,dx$$

for some constant $\lambda > 0$, then $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ with the least constant $K \sim N^{2\lambda}$ (see, e.g., [51, (7.13), (7.14)]). In such instances, Theorem 4.12 requires a considerably larger number of random points for effectively discretizing L_p norms.

Similar to the case of L_2-discretization, a change of density argument allows us to apply Theorem 4.12 to establish a weighted discretization theorem that does not require the $(2,\infty)$–Nikol'skii condition. The proof is similar to that of Corollary 3.8 for L_2-discretization. It relies on the following Lewis' change of density lemma:

Lemma 4.13 ([7, Lemma 7.1], [44]). *Given any $1 \leqslant p < \infty$ and any N-dimensional subspace X_N of $L_p(\Omega,\mu)$, there exists a basis $\{u_j\}_{j=1}^N$ of X_N such that the function $F := \left(\frac{1}{N}\sum_{j=1}^N |u_j|^2\right)^{\frac{1}{2}}$ has the following properties: $\|F\|_{L_p(\Omega,\mu)} = 1$, and for all scalars $\{\lambda_j\}_{j=1}^N \subset \mathbb{R}$,*

$$\int_\Omega \Big|\sum_{j=1}^N \lambda_j u_j(x)\Big|^2 F(x)^{p-2}\,d\mu(x) = \sum_{j=1}^N \lambda_j^2. \tag{4.21}$$

To illustrate the idea, let F be the function as given in Lemma 4.13. Consider the probability measure $\nu(dx) = |F(x)|^p\mu(dx)$ on Ω, and the linear mapping $U : L_p(\Omega,\mu) \to L_p(\Omega,\nu)$ given by

$$Uf(x) = \begin{cases} \frac{f(x)}{F(x)}, & \text{if } F(x) \neq 0; \\ 1, & \text{otherwise.} \end{cases}$$

Since each function in X_N is identically zero on the set $\{x\colon F(x) = 0\}$, we have

$$\|Uf\|_{L_p(\nu)} = \|f\|_{L_p(\mu)}, \quad \forall f \in X_N.$$

In addition, (4.21) implies that the sequence $\psi_j := Uu_j$, $j = 1,2,\dots,N$ forms an orthonormal basis of the space UX_N with respect to the inner product of $L_2(\Omega,d\nu)$. Moreover, we have

$$\sum_{j=1}^N |\psi_j(x)|^2 = N, \quad \forall x \in \Omega,$$

implying

$$\|Uf\|_{L_\infty(\Omega)} \leqslant N^{\frac{1}{2}} \|Uf\|_{L_2(\Omega,\nu)}, \quad \forall f \in X_N.$$

Thus, applying Theorem 4.12 to the space UX_N and the measure ν, we deduce the following result.

Corollary 4.14. *Let $1 \leqslant p < \infty$ and let X_N be an N-dimensional space of bounded functions on Ω associated with a function F given in Lemma 4.13. Consider a new measure ν given by*

$$d\nu = |F|^p \, d\mu.$$

Let $\{\xi^j\}_{j=1}^\infty$ be a sequence of IID random points drawn from the distribution ν. Let $\epsilon \in (0, 1/2)$. For $j = 1, 2, \dots, m$, let $\lambda_j := \frac{1}{m} \frac{1}{|F(\xi^j)|^p}$ if $F(\xi^j) \neq 0$, and $\lambda_j = \frac{1}{m}$ if $F(\xi^j) = 0$.

(i) *If $1 \leqslant p \leqslant 2$, then there exists a constant $C_p(\epsilon) \geqslant 1$ such that for any given parameter $A \geqslant 1$, and any integer*

$$m \geqslant C_p(\epsilon) A \cdot N \cdot \log^3(2N),$$

the Marcinkiewicz-type discretization inequalities

$$(1-\varepsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \sum_{j=1}^m \lambda_j \left|f\left(\xi^j\right)\right|^p \leqslant (1+\varepsilon)\|f\|^p_{L_p(\Omega,\mu)} \tag{4.22}$$

hold simultaneously for all $f \in X_N$ with probability $\geqslant 1 - N^{-A}$.

(ii) *If $2 < p < \infty$, then there exists a constant $C_p(\epsilon) \geqslant 1$ such that for any given parameter $A \geqslant 1$, and any integer*

$$m \geqslant C_p(\epsilon) A \cdot N^{p/2} \cdot \log^2(2N),$$

the Marcinkiewicz-type discretization inequality (4.22) holds simultaneously for all $f \in X_N$ with probability $\geqslant 1 - e^{-A}$.

5 Universal discretization

We begin with a general universal discretization result for a finite collection of linear spaces $\{V_j\}_{j=1}^L$ that satisfy an $\mathrm{NI}_{2,\infty}(\sqrt{R})$-condition with a common constant $R \geqslant 1$.

Theorem 5.1. *Let $\{V_j\}_{j=1}^L$ be a sequence of linear subspaces of $L_\infty(\Omega)$ which satisfies the following condition for some constant $R \geqslant 1$:*

$$\|f\|_{L_\infty(\Omega)} \leqslant \sqrt{R}\|f\|_{L_2(\Omega,\mu)}, \quad \forall f \in \bigcup_{j=1}^L V_j. \tag{5.1}$$

Let $\xi^1, \dots, \xi^m$ be a sequence of independent random variables taking values in Ω and satisfying

$$\frac{1}{m}\sum_{j=1}^m \mathbb{P}[\xi^j \in E] = \mu(E), \quad \forall E \in \mathcal{F}.$$

Let $\epsilon \in (0, \frac{1}{2})$ and $1 \leqslant p \leqslant 2$. Then there exist constants $C_p(\epsilon) \geqslant 1$ and $c_p > 0$ such that for any parameter $A \geqslant 1$ and any integer

$$m \geqslant C_p(\epsilon) \cdot A \cdot R\Big[(\log R)^3 + (\log R)^{3-\frac{p}{2}} (\log\log L)^{\frac{p}{2}} + (\log L)(\log R)\Big], \tag{5.2}$$

the Marcinkiewicz-type discretization inequality

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^{m}|f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}$$

holds simultaneously for all $f \in \bigcup_{j=1}^{L} V_j$ *with probability* $\geqslant 1 - R^{-A}$.

To the best of our knowledge, Theorem 5.1 is new for general linear subspaces of $L_\infty(\Omega)$. The proof proceeds similarly to that of Theorem 4.12. We summarize it below.

Proof. Note first that in the case where μ is the uniform probability measure on $\Omega = \{1, 2, \dots, n\}$ with $n \leqslant C_\epsilon R^2 \log L$, Theorem 5.1 follows directly from Part (i) of Corollary 4.8.

For a general probability measure μ, the proof of Theorem 5.1 proceeds in two steps. In the first step, we observe from (5.1) and (1.6) that $\dim V_j \leqslant R$ for each $1 \leqslant j \leqslant L$. Therefore, for $1 \leqslant p \leqslant 2$, applying Proposition 4.9 with $s = R$ yields a preliminary universal discretization using

$$m \approx C_\epsilon(R \log L + R^2)$$

random points. Although this sample size estimate is weaker than the bound stated in Theorem 5.1, it provides the crucial initial reduction step.

In the second step, we fix the random points obtained from the preliminary discretization and restrict all functions on this set. We then identify each V_j with a linear subspace of $\mathbb{R}^n$, where

$$n \approx R \log L + R^2.$$

Thus, Part (i) of Corollary 4.8 can be applied to further reduce the sample size. The technical details closely follow those in the proof of Theorem 4.12. □

We note that the sample size estimate (5.2) in Theorem 5.1 is optimal up to logarithmic factors when the number of linear spaces L is relatively small (e.g., $L \leqslant R^\alpha$ for some constant $\alpha \geqslant 1$). To illustrate this, consider the collection of N-dimensional subspaces of trigonometric polynomials $\{\mathcal{T}(Q_j)\}_{j=1}^{N^2}$ from Example 3.11, where $Q_j \subset \mathbb{Z}^d$ and $|Q_j| = N$. Theorem 5.1 guarantees universal discretization of the $L_p(\mathbb{T}^d)$ norm for $1 \leqslant p \leqslant 2$ and all $f \in \bigcup_{j=1}^{N^2} \mathcal{T}(Q_j)$ whenever the sample size satisfies $m \geqslant CN \log^3 N$.

On the other hand, the estimate in Theorem 5.1 can be sharpened if the collection $\{V_j\}_{j=1}^{L}$ has additional structure, such as being generated by a finite set of uniformly bounded functions on Ω, and L is substantially larger than R (e.g., $L \geqslant R^R$). For the remainder of this section, we focus precisely on this specific setting.

We introduce the general setting as follows. Let $\mathcal{D}_N = \{\varphi_i\}_{i=1}^{N}$ be a fixed system of linearly independent bounded functions on Ω satisfying

$$\sup_{x\in\Omega}|\varphi_j(x)| \leqslant 1, \quad j = 1, 2, \dots, N.$$

Given an integer $1 \leqslant s \leqslant N$, we denote by $\mathcal{X}_s(\mathcal{D}_N)$ the collection of all linear spaces spanned by $\{\varphi_j : j \in J\}$ with $J \subset \{1, 2, \dots, N\}$ and $|J| = s$. A function $f : \Omega \to \mathbb{C}$ is said to be s-sparse with

respect to $\mathcal{D}_N$ if it belongs to a linear space from the collection $\mathcal{X}_s(\mathcal{D}_N)$. We denote by $\Sigma_s(\mathcal{D}_N)$ the set of all s-sparse functions with respect to the dictionary $\mathcal{D}_N$; namely,

$$\Sigma_s(\mathcal{D}_N) := \bigcup_{V\in\mathcal{X}_s(\mathcal{D}_N)} V.$$

Recall that a finite sequence $\{\xi^j\}_{j=1}^m$ of points in Ω is said to provide universal discretization of the L_p norm for $\Sigma_s(\mathcal{D}_N)$, $1 \leqslant p < \infty$, with positive constants C_1, C_2 if

$$C_1\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant C_2\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in \Sigma_s(\mathcal{D}_N). \tag{5.3}$$

Consider a sequence $\{\xi^j\}_{j=1}^m$ of IID random points with common distribution μ. Our goal is to establish an estimate for the probability that (5.3) holds in terms of the number m of random points. In the above context, we have the following result, which improves the estimates in Theorem 5.1.

Theorem 5.2 ([21]). *Let $\mathcal{D}_N = \{\varphi_j\}_{j=1}^N$ be a set of N bounded functions on Ω satisfying*

$$\max_{1\leqslant j\leqslant N} \|\varphi_j\|_{L_\infty(\Omega)} \leqslant 1.$$

Let $1 \leqslant s \leqslant N$ be a given integer. Assume that there exists a constant $K \geqslant 1$ such that

$$\sum_{j\in J} |a_j|^2 \leqslant K\Big\|\sum_{j\in J} a_j\varphi_j\Big\|^2_{L_2(\Omega,\mu)}, \quad \forall a_j \in \mathbb{C} \tag{5.4}$$

whenever $J \subset \{1, 2, \dots, N\}$ with $|J| = s$. Let $\xi^1, \dots, \xi^m$ be IID random points with common distribution μ on Ω. Then given $1 \leqslant p \leqslant 2$ and $\epsilon \in (0, \frac{1}{2})$, there exist constants $C_p(\varepsilon) > 1$ and $c_p(\varepsilon) > 0$, depending only on p and ε, such that the inequalities

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in \Sigma_s(\mathcal{D}_N)$$

hold with probability $\geqslant 1 - 2\exp\Big(-\frac{c_p(\epsilon)m}{Ks\log^2(2Ks)}\Big)$ provided that

$$m \geqslant C_p(\epsilon)Ks\log N \cdot (\log(2Ks))^2 \cdot (\log(2Ks) + \log\log N).$$

Theorem 5.2 was proved in [21]. Note that the problem of universal discretization for some subspaces of the trigonometric polynomials was studied in [18, 63].

The condition (5.4) is satisfied for all $1 \leqslant s \leqslant N$ if $\varphi_1, \dots, \varphi_N$ are linearly independent in $L_2(\Omega, \mu)$, in which case one can choose K^{-1} to be the smallest eigenvalue of the $N \times N$ matrix $\Big[\langle\varphi_j, \varphi_k\rangle_{L_2(\Omega,\mu)}\Big]_{1\leqslant j,k\leqslant N}$. In applications, we often assume that $\mathcal{D}_N$ is a Riesz basis; that is, there exist constants $0 < R_1 \leqslant R_2 < \infty$ such that for any $(a_1, \dots, a_N) \in \mathbb{C}^N$,

$$R_1\Big(\sum_{j=1}^N |a_j|^2\Big)^{1/2} \leqslant \Big\|\sum_{j=1}^N a_j\varphi_j\Big\|_{L_2(\Omega,\mu)} \leqslant R_2\Big(\sum_{j=1}^N |a_j|^2\Big)^{1/2}.$$

Note however that estimates in Theorem 5.2 are independent of the constant R_2.

The universal discretization of L_2 norm is closely related to the concept of Restricted Isometry Property (RIP), which plays a central role in compressed sensing (see, e.g., [28] and the references therein). Given an integer $1 \leqslant s \leqslant N$, we denote by $\Sigma_s(\mathbb{R}^N)$ the set of all s-sparse vectors in $\mathbb{R}^N$; i.e., vectors $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$ with at most s nonzero coordinates x_i. An $m \times N$ matrix $\mathbf{A}$ is said to satisfy the RIP of order s if there exists a positive constant $\delta \in (0,1)$ such that

$$(1-\delta)\|\mathbf{x}\|^2_{\ell_2^N} \leqslant \|\mathbf{A}\mathbf{x}\|^2_{\ell_2^m} \leqslant (1+\delta)\|\mathbf{x}\|^2_{\ell_2^N}, \quad \forall \mathbf{x} \in \Sigma_s(\mathbb{R}^N). \tag{5.5}$$

The smallest positive constant δ that satisfies (5.5), denoted by $\delta_s(\mathbf{A})$, is called the restricted isometry constant.

Results on the RIP properties of the random matrix $\mathbf{\Phi}(\boldsymbol{\xi})$ associated with uniformly bounded orthonormal systems $\mathcal{D}_N$ can be found in [28]. Generalizations of the RIP, where one or both of the ℓ_2 norms are replaced by other norms, typically ℓ_p^N norms, have also proven useful in this context (see, e.g., [3, 10], [28, Exercise 9.6]). To express universal sampling discretization within the RIP framework, the following extension was introduced in [64, Section 5].

Definition 5.3. Let $\|\cdot\|$ be a norm on $\mathbb{R}^N$, and let $1 \leqslant p < \infty$. Let $1 \leqslant s, m \leqslant N$ be integers. An $m \times N$ real matrix $\mathbf{A}$ is said to have the s-th order $RIP\left(\ell_p^m, \|\cdot\|\right)$ with constants $C_1, C_2 > 0$ if for any $\mathbf{x} \in \Sigma_s(\mathbb{R}^N)$, we have

$$C_1\|\mathbf{x}\|^p \leqslant \|\mathbf{A}\mathbf{x}\|^p_{\ell_p^m} \leqslant C_2\|\mathbf{x}\|^p.$$

Next, let $\Phi : \Omega \to \mathbb{R}^N$ be a vector-valued function on Ω given by

$$\Phi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \dots, \varphi_N(\mathbf{x})), \quad \mathbf{x} \in \Omega.$$

Here and elsewhere in this section, we always treat $\Phi(\mathbf{x})$ as a row vector. Given a finite sequence $\boldsymbol{\xi} := \{\xi^j\}_{j=1}^m \subset \Omega$ of points, consider the $m \times N$ random matrix

$$\mathbf{\Phi}(\boldsymbol{\xi}) := \begin{bmatrix} \Phi(\xi^1) \\ \Phi(\xi^2) \\ \vdots \\ \Phi(\xi^m) \end{bmatrix} = \begin{bmatrix} \varphi_1(\xi^1) & \varphi_2(\xi^1) & \cdots & \varphi_N(\xi^1) \\ \varphi_1(\xi^2) & \varphi_2(\xi^2) & \cdots & \varphi_N(\xi^2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(\xi^m) & \varphi_2(\xi^m) & \cdots & \varphi_N(\xi^m) \end{bmatrix}. \tag{5.6}$$

Additionally, given $1 \leqslant p < \infty$, we define the norm $\|\cdot\|_{p,\Phi}$ on $\mathbb{R}^N$ by

$$\|\mathbf{a}\|_{p,\Phi} = \Big\| \sum_{j=1}^N a_j \varphi_j \Big\|_{L_p(\Omega,\mu)}, \quad \mathbf{a} = (a_1, \dots, a_N)^T \in \mathbb{R}^N.$$

With the above notation, we can then write the inequality (5.3) in the form

$$C_1\|\mathbf{a}\|_{p,\Phi} \leqslant \big\|\mathbf{\Phi}(\boldsymbol{\xi})\mathbf{a}\big\|_{L_p^m} \leqslant C_2\|\mathbf{a}\|_{p,\Phi}, \quad \forall \mathbf{a} \in \Sigma_s(\mathbb{R}^N),$$

which is equivalent to asserting that matrix $m^{-1/p}\mathbf{\Phi}(\boldsymbol{\xi})$ has the s-th order $RIP\left(\ell_p^m, \|\cdot\|_{p,\Phi}\right)$ with constants $C_1, C_2 > 0$. In particular, if $p = 2$, $\epsilon \in (0,1)$ and $\varphi_1, \dots, \varphi_N$ is an orthonormal system in $L_2(\Omega, \mu)$, then the universal discretization (5.3) with $C_1 = 1 - \epsilon$ and $C_2 = 1 + \epsilon$ can be formulated equivalently in terms of the RIP property of order s of the random matrix $m^{-1/2}\mathbf{\Phi}(\boldsymbol{\xi})$:

$$(1-\epsilon)\|z\|^2_{\ell_2^N} \leqslant m^{-1/2}\|\mathbf{\Phi}(\boldsymbol{\xi})z\|_{\ell_2^m} \leqslant (1+\epsilon)\|z\|^2_{\ell_2^N}, \quad \forall z \in \Sigma_s(\mathbb{R}^N).$$

In summary, we state the following result.

Theorem 5.4. *Assume that $\mathcal{D}_N = \{\varphi_1, \dots, \varphi_N\}$ is a finite system of real-valued functions on Ω, $1 \leqslant p < \infty$ and $1 \leqslant s, m \leqslant N$ are integers. Then a sequence $\boldsymbol{\xi} := \{\xi^j\}_{j=1}^m$ of m points in Ω provides the universal discretization* (5.3) *of the L_p norm for $\Sigma_s(\mathcal{D}_N)$ with constants $C_1, C_2 > 0$ if and only if the $m \times N$ matrix $m^{-1/p}\boldsymbol{\Phi}(\boldsymbol{\xi})$, with $\boldsymbol{\Phi}(\boldsymbol{\xi})$ being defined in* (5.6)*, has the s-th order RIP $\left(\ell_p^m, \|\cdot\|_{p,\Phi}\right)$ with constants $C_1, C_2 > 0$. In particular, if $p = 2$, $\epsilon \in (0,1)$ and $\mathcal{D}_N$ is an orthonormal system in $L_2(\Omega, \mu)$, then the universal discretization* (5.3) *with constants $C_1 = 1 - \epsilon$ and $C_2 = 1 + \epsilon$ holds if and only if the matrix $m^{-1/2}\boldsymbol{\Phi}(\boldsymbol{\xi})$ has the RIP property of order s with constants $1 \pm \epsilon$.*

Some applications of universal discretization and Theorem 5.4 can be found in [23, 22].

6 Some improved bounds in sampling discretization of integral norms

Analogous to the L_2-discretization discussed in Subsection 3.3, for any given constants $C_2 \geqslant C_1 > 0$ and $1 \leqslant p < \infty$, we define $m(X_N; p; C_1, C_2)$ to be the smallest positive integer m for which there exist m points $\xi^1, \dots, \xi^m \in \Omega$ such that

$$C_1 \|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m} \sum_{j=1}^m |f(\xi^j)|^p \leqslant C_2 \|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in X_N.$$

For simplicity, we denote this quantity by $m(X_N; p)$ when the specific values of the absolute constants C_1, C_2 are understood or not important from context.

According to Theorem 4.12, for every space X_N satisfying the $(2,\infty)$-Nikol'skii inequality $\mathrm{NI}_{2,\infty}(\sqrt{KN})$, we have

$$m(X_N; p) \leqslant C_p \begin{cases} KN \log^3(KN), & \text{if } 1 \leqslant p \leqslant 2; \\ (KN)^{p/2} \log^2(KN), & \text{if } 2 < p < \infty. \end{cases} \tag{6.1}$$

The upper bounds in (6.1) turn out to be nearly optimal as $N \to \infty$ for a general space $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$. Indeed, this is clear for $1 \leqslant p \leqslant 2$ as the lower bound $m(X_N; p) \geqslant N$ holds trivially. For $2 < p < \infty$, it is known (see, for instance, [32, **D.20**]) that for $X_N = \mathrm{span}\{r_1, \dots, r_N\}$ with $r_j(t) := \mathrm{sgn}(\sin(2^{j+1}\pi t))$, $t \in [0,1]$, $j \in \mathbb{N}$ being the sequence of Rademacher functions on $[0,1]$ and for the ususal Lebesgue measure on $\Omega = [0,1]$, the $(2,\infty)$-Nikol'skii inequality $\mathrm{NI}_{2,\infty}(\sqrt{KN})$ is satisfied with $K = 1$, and $m(X_N; p) \geqslant cN^{p/2}$ for any $2 < p < \infty$, where $c > 0$ is an absolute constant.

It is natural to ask whether the upper bounds in (6.1) remain valid without the extra logarithmic factors under the condition $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$. For the discretization of L_2 norm, the answer is positive, due to the breakthrough work of A. Marcus, D. Spielman, and N. Srivastava [50]. In fact, according to Theorem 3.9, we have

$$m(X_N; 2; 1/2, 3/2) \leqslant CKN.$$

For $1 \leqslant p < 2$, this remains a challenging open problem.

Recent progress on this problem for $1 \leqslant p < 2$ was established in [14]. Specifically, for $1 \leqslant p < 2$ and every $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$, the refined upper bound $m(X_N; p; 1/2, 3/2) \leqslant C_p \Phi_p(KN)$ holds with

$$\Phi_p(t) := \begin{cases} t \log t, & \text{if } p = 1,\ 2, \\ t(\log t)(\log\log t)^2, & \text{if } 1 < p < 2, \end{cases} \quad t \geqslant 2. \tag{6.2}$$

Theorem 6.1 ([14]). *Given $1 \leqslant p \leqslant 2$, $0 < \epsilon \leqslant \frac{1}{2}$, $K \geqslant 2$ and any N-dimensional subspace $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$, there exist $\xi^1, \dots, \xi^m \in \Omega$ with $m \leqslant C_p(\epsilon)\Phi_p(KN)$ such that*

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)}, \quad \forall f \in X_N, \tag{6.3}$$

where Φ_p is given in (6.2), $C_p(\epsilon) = C\epsilon^{-2}$ for $p = 1, 2$, and $C_p(\epsilon) = C_p\epsilon^{-2}\log^3\frac{1}{\epsilon}$ for $1 < p < 2$.

Historical discussions on sampling discretization for the cases $p = 2$ and $1 \leqslant p < 2$ can be found in Subsections **D.15** and **D.16** of [32], respectively. Here we only mention that the best previously known results are as follows. It was proved in [16] that, under the $(2,\infty)$–Nikol'skii inequality assumption $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$, the discretization estimate (6.3) holds for $1 \leqslant p < 2$ provided that $m \geqslant C(p, K, \varepsilon)\, N(\log N)^3$. This estimate on m was improved to $m \geqslant C(p, K, \varepsilon)N(\log N)^2$ for $1 < p < 2$ and to $m \geqslant C(p, K, \varepsilon)N^{p/2}\log N$ for $p > 2$ in [36]. Furthermore, sampling discretization for $p > 2$ under the (p,∞)-Nikol'skii inequality $X_N \in \mathrm{NI}_{p,\infty}((KN)^{1/p})$ was studied in [36], with subsequent improvements for $p > 3$ presented in [19]. More recently, [13] improved the sample size estimates from [36, 19] under the (p,∞)-Nikol'skii inequality for $p > 2$, proving that the discretization inequality (6.3) holds whenever $m \geqslant C(p, K, \varepsilon)N(\log N)(\log\log N)^{p-1}$.

Combining Theorem 6.1 with Lewis' change of density lemma (see Lemma 4.13), and following the proof of Corollary 4.14, we obtain the following weighted discretization result.

Corollary 6.2. *Given $1 \leqslant p \leqslant 2$, $0 < \epsilon \leqslant \frac{1}{2}$, and any N-dimensional subspace $X_N \subset L_p(\Omega,\mu)$, there are a finite set of points $\{\xi^1, \dots, \xi^m\} \subset \Omega$ with $m \leqslant C_p(\epsilon)\Phi_p(KN)$ and a set of nonnegative weights $\{\lambda_j\}_{j=1}^m$ such that for any $f \in X_N$,*

$$(1-\epsilon)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \sum_{j=1}^m \lambda_j |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|^p_{L_p(\Omega,\mu)},$$

where Φ_p is given in (6.2), $C_p(\epsilon) = C\epsilon^{-2}$ for $p = 1, 2$, and $C_p(\epsilon) = C_p\epsilon^{-2}\log^3\frac{1}{\epsilon}$ for $1 < p < 2$.

We give a brief description of the scheme of the proof of Theorem 6.1, which can be divided into three steps.

Step 1. Preliminary discretization.

In this step, we establish results on simultaneous discretization of the L_2 and L_p norms, using $M \leqslant C_p\varepsilon^{-r_1}N(\log N)^{r_2}$ points. This step allows us to reduce the original problem of discretization to a problem in $\mathbb{R}^M$. The main ingredient in this step is the following lemma, which follows directly from Proposition 4.9.

Lemma 6.3. *Let $1 \leqslant p < 2$ and $0 < \epsilon_0 < 1/4$. Let X_N be a subspace of $L_\infty(\Omega)$ of dimension N such that $X_N \in \mathrm{NI}_{2,\infty}(\sqrt{KN})$ for some $K \geqslant 2$. Then there exists a finite set of points $x_1, \dots, x_m \in \Omega$ with $m \leqslant C_p\epsilon_0^{-8}KN(\log(KN))^3$ such that for any $f \in X_N$, we have*

$$(1-\epsilon_0)\|f\|^p_{L_p(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(x_j)|^p \leqslant (1+\epsilon_0)\|f\|^p_{L_p(\Omega,\mu)},$$

and

$$(1-\epsilon_0)\|f\|^2_{L_2(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(x_j)|^2 \leqslant (1+\epsilon_0)\|f\|^2_{L_2(\Omega,\mu)}.$$

Step 2. Estimate the expectation $\mathbb{E}\Big(\sup_{f\in X_N\cap B_p^M}\Big|\sum_{j=1}^M \varepsilon_j|f(j)|^p\Big|\Big)$, where $\{\varepsilon_j\colon j=1,2,\dots,M\}$ denotes a sequence of independent Bernoulli random variables taking values ±1 with probability $1/2$, and $B_p^M := \{f\in\mathbb{R}^M\colon \|f\|_{\ell_p^M}\leqslant 1\}$.

For $p=1,2$, we have the following lemma, which was proved by Rudelson [56, Lemma 1] for $p=2$, and by Talagrand [58] for $p=1$ (see also [14, Theorem 3.1], [30, Theorem 13] and [43, Proposition 15.16]).

Lemma 6.4. *Let X_N be a subspace of $\mathbb{R}^M$ of dimension at most N satisfying*

$$\|f\|_\infty \leqslant \sqrt{KN}\|f\|_{L_2^M}, \qquad \forall f\in X_N$$

for a constant $K\geqslant 1$. Then for $p=1$ and $p=2$ we have

$$\mathbb{E}\Big(\sup_{f\in X_N\cap B_p^M}\Big|\sum_{j=1}^M \varepsilon_j|f(j)|^p\Big|\Big)\leqslant C\sqrt{\frac{KN\log N}{M}},$$

where C is a positive absolute constant.

The subsequent result for $1<p<2$ builds upon the approach established in Talagrand's proof of [59, Theorem 16.8.2].

Lemma 6.5 ([14, Theorem 4.2])**.** *Let X_N be a subspace of $\mathbb{R}^M$ of dimension at most N satisfying*

$$\|f\|_\infty \leqslant \sqrt{KN}\|f\|_{L_2^M}, \qquad \forall f\in X_N$$

for some constant $K\geqslant 1$. Then for any $p\in(1,2)$, we have

$$\mathbb{E}\Big(\sup_{f\in X_N\cap B_p^M}\Big|\sum_{j=1}^M \varepsilon_j|f(j)|^p\Big|\Big)\leqslant C(p)\sqrt{\frac{KN\log M}{M}}\log\Big(\frac{M}{KN}+2\Big).$$

For $1<p<2$, Talagrand [59, Theorem 16.8.2] proved a result similar to Lemma 6.5 for a probability measure μ on $\mathbb{R}^M$ satisfying $\mu\{j\}\leqslant\frac{2}{M}$ for $1\leqslant j\leqslant M$, under the following stronger assumption: for each orthonormal basis $\{\varphi_k\}_{k=1}^N$ of $(X_N,\|\cdot\|_{L_2(\mu)})$,

$$\frac{1}{N}\sum_{k=1}^N \varphi_k(j)^2 = 1, \quad j=1,2,\dots,M.$$

Lemmas 6.4 and 6.5 play crucial roles in the proof of Theorem 6.1. Roughly speaking, they allow us to reduce the number of points required for good discretization by approximately a factor of two.

Step 3. Iteration.

To illustrate the idea, we sketch the proof of $p=1$ below.

First, by Lemma 6.3 in the preliminary step, without loss of generality, we may assume that $\Omega = \Omega_M = \{1,2,\dots,M\}$ and μ is the uniform probability measure on Ω_M. For each $I\subset\Omega_M$, we denote by R_I the orthogonal projection onto the space spanned by $\{e_i, i\in I\}$, where $e_1=(1,0,\dots,0)$, …, $e_M=(0,\dots,0,1)$ is a canonical basis of $\mathbb{R}^M$. Thus, for each $f\in\mathbb{R}^M$, $(R_If)(j)=f(j)$ for $j\in I$, and $(R_If)(j)=0$ for $j\in\Omega_M\backslash I$. Furthermore, an application of Lemma 6.4 yields the following structural result.

Lemma 6.6 ([14, Lemma 3.2]). *Let X_N be a subspace of $\mathbb{R}^M$ of dimension at most N satisfying*

$$\|f\|_\infty \leqslant \sqrt{KN}\|f\|_{L_2^M}, \quad \forall f \in X_N,$$

for some constant $K \geqslant 1$. Let $J \subset \Omega_M := \{1, 2, \dots, M\}$. Assume that there exist positive constants α_J, β_J such that for any $f \in X_N$ we have for both $p = 1$ and $p = 2$

$$\alpha_J \|f\|_{\ell_p^M}^p \leqslant \|R_J f\|_{\ell_p^M}^p \leqslant \beta_J \|f\|_{\ell_p^M}^p.$$

Then there exists a subset $I \subset J$ with

$$\frac{|J|}{2}\Big(1 - \frac{1}{\sqrt{|J|}}\Big) \leqslant |I| \leqslant \frac{|J|}{2}$$

such that for any $f \in X_N$ we have for both $p = 1$ and $p = 2$,

$$\alpha_I \|f\|_{\ell_p^M}^p \leqslant \|R_I f\|_{\ell_p^M}^p \leqslant \beta_I \|f\|_{\ell_p^M}^p$$

where

$$\alpha_I := \frac{(1-\sigma)\alpha_J}{2}, \quad \beta_I := \frac{(1+\sigma)\beta_J}{2}, \quad \sigma := C\sqrt{\frac{KN\log N}{\alpha_J M}},$$

and C is an absolute constant.

We can iterate Lemma 6.6 with an appropriate stopping time and finish the proof.

Example 6.7. Let $Q \subset \mathbb{Z}^d$ be an arbitrary set of frequencies, and consider the space

$$\mathcal{T}(Q) := \operatorname{span}\big\{e^{\mathrm{i}\langle k, \cdot\rangle} : k \in Q\big\}$$

of all trigonometric polynomials with frequencies from Q on the cube $\Omega = [0, 2\pi)^d$, equipped with the normalized Lebesgue measure. Then $\mathcal{T}(Q) \in \mathrm{NI}_{2,\infty}(\sqrt{N})$ with $N := \dim \mathcal{T}(Q) = |Q|$ and Theorem 6.1 implies that there exist $m \leqslant C\epsilon^{-2} N \log N$ points $\xi^1, \dots, \xi^m \in \Omega$ such that

$$(1-\epsilon)\|f\|_{L_1(\Omega,\mu)} \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)| \leqslant (1+\epsilon)\|f\|_{L_1(\Omega,\mu)}, \quad \forall f \in \mathcal{T}(Q).$$

Similarly, for $1 < p < 2$, there exist $m \leqslant C_p(\varepsilon) N \log N (\log\log N)^2$ points $\xi^1, \dots, \xi^m \in \Omega$ such that

$$(1-\epsilon)\|f\|_{L_p(\Omega,\mu)}^p \leqslant \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \leqslant (1+\epsilon)\|f\|_{L_p(\Omega,\mu)}^p, \quad \forall f \in \mathcal{T}(Q).$$

7 Sampling discretization in the uniform norm

In this section, we will survey some recent progress on sampling discretization of the uniform norm. We assume that Ω is a compact subset of $\mathbb{R}^d$ equipped with a Borel probability measure μ. Let X_N denote an N-dimensional subspace of the space $\mathcal{C}(\Omega)$ of continuous functions on Ω. For $A \subset \Omega$ and $f \in \mathcal{C}(\Omega)$, define $\|f\|_A = \sup_{x\in A} |f(x)|$. We say that X_N admits an L_∞- Bernstein-type

discretization theorem with parameter $m \in \mathbb{N}$ and positive constant C if there exist m points $\xi^j \in \Omega$, $j = 1, 2, \ldots, m$, for which

$$\|f\|_\Omega \leqslant C \max_{1 \leqslant j \leqslant m} \left|f\left(\xi^j\right)\right|, \quad \forall f \in X_N. \tag{7.1}$$

For convenience, we denote by $m(X_N; \infty; C)$ the minimum number m of points required for the discretization (7.1) with a given constant $C > 0$.

The first results on discretization of the uniform norm were obtained by Bernstein [4, 5] (see also [68], Ch.10, Theorem (7.28)). In recent years this problem has been extensively studied in various settings (e.g. see [15, 31, 33, 38]). We will focus on recent progress on sampling discretization of L_∞ norm. We refer to the surveys [18, 32] and the references therein for historical comments on this problem.

7.1 General bounds of $m(X_N; \infty; C)$ with C being independent of N

We start with the following general well known result (see, e.g., [31, Theorem 1.3]), which in particular gives the exponential upper bound $m(X_N; \infty; 2) \leqslant 9^N$.

Theorem 7.1. [1] *Let X_N be an N-dimensional subspace of $L_\infty(\Omega)$. Given any $\epsilon \in (0, 1]$, there exists a sequence $\left\{\xi^j\right\}_{j=1}^m$ of $m \leqslant \left(1 + \frac{8}{\epsilon}\right)^N$ points in Ω satisfying*

$$\max_{x \in \Omega} |f(x)| \leqslant (1 + \epsilon) \max_{1 \leqslant j \leqslant m} \left|f\left(\xi^j\right)\right|, \quad \forall f \in X_N. \tag{7.2}$$

Proof. Using (4.2), we can find a finite sequence

$$\{f_j\}_{j=1}^m \subset B := \{f \in X_N : \ \|f\|_{L_\infty(\Omega)} \leqslant 1\} \ \text{ with } \ m \leqslant \left(1 + \frac{8}{\epsilon}\right)^N$$

such that $B \subset \bigcup_{j=1}^m (f_j + \frac{\epsilon}{4} B)$. For each $1 \leqslant j \leqslant m$, let $\xi^j \in \Omega$ be such that $|f_j(\xi^j)| = \|f_j\|_{L_\infty(\Omega)}$. For any $f \in X_N$ with $\|f\|_{L_\infty(\Omega)} = 1$, we can find an integer $1 \leqslant j \leqslant m$ such that $\|f_j - f\|_{L_\infty(\Omega)} \leqslant \frac{\epsilon}{4}$, which implies

$$\|f\|_{L_\infty(\Omega)} \leqslant \|f_j\|_{L_\infty(\Omega)} + \frac{\epsilon}{4} \leqslant |f(\xi^j)| + \frac{\epsilon}{2} \|f\|_{L_\infty(\Omega)}.$$

The estimate (7.2) then follows since $(1 - \frac{\epsilon}{2})^{-1} \leqslant 1 + \epsilon$. □

Theorem 7.1 establishes an exponential upper bound in the dimension N for the number of points required to discretize the uniform norm in X_N. In general, the discretization (7.1) of the uniform norm in X_N with a constant C independent of N may not be possible unless the number m of required points grows exponentially with N (see, e.g., [33] and [31, Theorem 1.2]). More precisely, the following result (see [33]), which complements Theorem 7.1, shows that there exists an N-dimensional subspace $X_N \subset C[0, 2\pi]$ for which

$$m(X_N; \infty; C) \geqslant (N/e) e^{cN/C^2} \ \text{ with } c > 0 \text{ being an absolute constant.}$$

[1] One of the referees kindly pointed out that the bound in this theorem can be improved from $(1 + \frac{8}{\epsilon})^N$ to $(1 + \frac{4}{\epsilon})^N$ by working with an $(\epsilon/2)$-net for the set of the extreme points of the unit ball of the dual space.

Theorem 7.2 ([33]). *Let $X_N := \operatorname{span}\{e^{ik_j x} : j = 1, 2, \dots, N\}$, where $\{k_j\}_{j=1}^N$ is a lacunary sequence of positive integers satisfying that $k_1 = 1$ and $k_{j+1} \geqslant bk_j$, $j = 1, 2, \dots, N-1$ for some constant $b > 1$. If there exist a sequence $\{\xi^j\}_{j=1}^m \subset [0, 2\pi)$ of m points and a constant $C > 0$ such that*

$$\max_{x\in[0,2\pi]} |f(x)| \leqslant C \max_j \left|f\left(\xi^j\right)\right|, \quad \forall f \in X_N,$$

then one must have

$$m \geqslant (N/e)e^{cN/C^2}$$

with a positive constant c which may only depend on b.

7.2 Estimates of $m(X_N; \infty; C_N)$ with C_N being dependent on N

Theorems 7.1 and 7.2 provide upper and lower estimates of the number of sample points needed for the discretization inequality (7.1) with C being independent of N. We now mention some results, where the discretization constant C in (7.1) is allowed to depend on N. In the classical setting of univariate algebraic and trigonometric polynomials, this problem was extensively studied (see, e.g., [11, 27]). However, as is customary in this survey, our goal is to consider a more general framework of abstract finite-dimensional spaces.

We start with the following result, which is known as the Auerbach lemma when $\ell = 1$ (see [54, Proposition 1.2.3]).

Theorem 7.3. *Let X_N be an N-dimensional real subspace of $\mathcal{C}(\Omega)$. Given a positive integer ℓ, let*

$$X_N(\ell) := \operatorname{span}\Big\{ f_1 f_2 \cdots f_\ell : \quad f_1, \dots, f_\ell \in X_N \Big\},$$

and $N_\ell := \dim X_N(\ell)$. Then there exists a set $\{\xi^j\}_{j=1}^{N_\ell}$ of points in Ω such that

$$\max_{x\in\Omega} |f(x)| \leqslant (N_\ell)^{\frac{1}{\ell}} \max_{1\leqslant j\leqslant N_\ell} \left|f\left(\xi^j\right)\right|, \quad \forall f \in X_N.$$

Proof. Given a basis $\{\varphi_1, \dots, \varphi_{N_\ell}\} \subset X_N(\ell)$ of $X_N(\ell)$, we define a function $\Delta : \Omega^{N_\ell} \to \mathbb{R}$ by

$$\Delta(x_1, x_2, \dots, x_{N_\ell}) := \det \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_{N_\ell}(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_{N_\ell}(x_2) \\ \vdots & \vdots & \cdots & \vdots \\ \varphi_1(x_{N_\ell}) & \varphi_2(x_{N_\ell}) & \cdots & \varphi_{N_\ell}(x_{N_\ell}) \end{bmatrix}$$

for $x_1, \dots, x_{N_\ell} \in \Omega$. Since Δ is a continuous function on the compact set Ω^{N_ℓ}, there must exist $\xi^1, \dots, \xi^{N_\ell} \in \Omega$ such that

$$\Delta(\xi^1, \dots, \xi^{N_\ell}) \equiv \Delta_{N,\ell} := \max_{x_1,\dots,x_{N_\ell}\in\Omega} \Delta(x_1, \dots, x_{N_\ell}).$$

It is easily seen that $\Delta_{N,\ell} > 0$. For $1 \leqslant j \leqslant N_\ell$, define $L_j : \Omega \to \mathbb{R}$ by

$$L_j(x) = \frac{\Delta(\xi^1, \dots, \xi^{j-1}, x, \xi^{j+1}, \dots, \xi^{N_\ell})}{\Delta_{N,\ell}}, \quad x \in \Omega.$$

Clearly, $L_j \in X_N(\ell)$, $L_j(\xi^k) = \delta_{k,j}$ and $\|L_j\|_{L_\infty(\Omega)} = 1$ for $1 \leqslant j, k \leqslant N_\ell$. Moreover, for each $f \in X_N$, we have

$$\big(f(x)\big)^\ell = \sum_{j=1}^{N_\ell} f(\xi^j)^\ell L_j(x), \quad x \in \Omega.$$

This implies that for each $f \in X_N$,

$$\|f\|^\ell_{L_\infty(\Omega)} \leqslant N_\ell \max_{1\leqslant j\leqslant N_\ell} |f(\xi^j)|^\ell.$$

Hence,

$$\|f\|_{L_\infty(\Omega)} \leqslant (N_\ell)^{\frac{1}{\ell}} \max_{1\leqslant j\leqslant N_\ell} |f(\xi^j)|, \quad \forall f \in X_N.$$

This completes the proof. □

We give a few remarks on Theorem 7.3.

Remark 7.4. (i) Of particular interest is the case when $\ell = 1$, where Theorem 7.3 gives the discretization inequality with the minimum number of points:

$$\max_{x\in\Omega} |f(x)| \leqslant N \max_{1\leqslant j\leqslant N} \left|f\left(\xi^j\right)\right|, \quad \forall f \in X_N. \tag{7.3}$$

Using the notation introduced in the beginning of this section, this implies $m(X_N; \infty; N) = N$. This is a well known result, for which the above proof can be found in the book of E. Novak [54, Proposition 1.2.3].

(ii) For any positive integer ℓ, we have

$$N_\ell \leqslant \binom{N-1+\ell}{\ell} \leqslant \Big(\frac{e(N-1+\ell)}{\ell}\Big)^\ell.$$

However, in many cases, this last estimate is far from being optimal. For example, if $X_N = \Pi^d_n$ is the space of all algebraic polynomials of total degree at most n in d variables, and Ω is a compact subset of $\mathbb{R}^d$ with nonempty interior, then

$$N = \dim \Pi^d_n = \binom{n+d}{d} \quad \text{and} \quad N_\ell \leqslant \dim \Pi^d_{n\ell}.$$

As a result, we have

$$\frac{N_\ell}{N} \leqslant \frac{\binom{n\ell+d}{d}}{\binom{n+d}{d}} = \frac{(n\ell+d)(n\ell+d-1)\cdots(n\ell+1)}{(n+d)(n+d-1)\cdots(n+1)} = \ell^d \prod_{j=1}^{d} \frac{n+\frac{j}{\ell}}{n+j} \leqslant \ell^d.$$

Assume in addition that $n \geqslant d^2$ and $d \geqslant 2$. Then

$$N = \prod_{j=1}^{d}\Big(1+\frac{n}{j}\Big) \geqslant \Big(1+\frac{n}{d}\Big)^d \geqslant d^d.$$

Now we specify the integer $\ell \in \mathbb{N}$ such that

$$\log N - 1 \leqslant \ell < \log N.$$

Then $\ell \geqslant d\log d - 1$,

$$N_\ell \leqslant \ell^d N \leqslant N(\log N)^d,$$

and

$$N_\ell^{\frac{1}{\ell}} \leqslant (\ell^d N)^{\frac{1}{\ell}} = \exp\Big(\frac{d}{\ell}\log\ell + \frac{1}{\ell}\log N\Big) \leqslant C_1,$$

where $C_1 > 0$ is an absolute constant (independent of d). Thus, Theorem 7.3 provides a set $\{\xi^j\}_{j=1}^m$ of

$$m \leqslant N_\ell \leqslant N(\log N)^d$$

points in Ω such that

$$\max_{x\in\Omega}|f(x)| \leqslant C_1 \max_{1\leqslant j\leqslant m}|f(\xi^j)|, \quad \forall f \in X_N = \Pi_n^d.$$

This last estimate was previously established in [6, Proposition 23], with implicit constants depending on d.[2]

If the number of sampling points m is relaxed to be of order N rather than exactly equal to N, one can prove a refinement of the discretization inequality (7.3), in which the constant N is improved to $C\sqrt{N}$. We state the following result, which follows directly from the proof of Theorem 2 of [39].

Theorem 7.5 ([39, Theorem 2]). *Given any real N-dimensional subspace X_N of $\mathcal{C}(\Omega)$ and any $\epsilon \in (0,1]$, there exists a set $\{\xi^j\}_{j=1}^m \subset \Omega$ of $m \leqslant (1+\epsilon)(N+1)$ points such that*

$$\max_{x\in\Omega}|f(x)| \leqslant C\sqrt{N}\epsilon^{-1}\max_{1\leqslant j\leqslant m}|f(\xi^j)|,$$

where $C > 0$ is an absolute constant.

For completeness, we include the proof from [39] below. The argument relies on Theorem 3.12 and the following result by J. Kiefer and J. Wolfowitz [34].

Lemma 7.6 ([34]). *Given a real N-dimensional subspace X_N of $\mathcal{C}(\Omega)$, there exists a probability measure μ on Ω such that for any $f \in X_N$,*

$$\max_{x\in\Omega}|f(x)| \leqslant \sqrt{N}\|f\|_{L_2(\mu)}.$$

Proof of Theorem 7.5. Let $X_{N+1} = \operatorname{span}\big(X_N \cup \{1\}\big)$. Clearly, $\dim X_{N+1} \leqslant N+1$. Applying Lemma 7.6 to the space X_{N+1} instead of X_N, we can find a probability measure μ on Ω for which

$$\max_{x\in\Omega}|f(x)| \leqslant \sqrt{N+1}\|f\|_{L_2(\mu)}, \quad \forall f \in X_{N+1}.$$

By Theorem 3.12 applied to $b = 1+\epsilon$ and the space $X_{N+1} \subset L_2(\mu)$, there exist a set $\{\xi^j\}_{j=1}^m \subset \Omega$ of $m \leqslant (1+\epsilon)(N+1)$ points and a sequence $\{\lambda_j\}_{j=1}^m$ of nonnegative weights such that

$$\|f\|_{L_2(\mu)} \leqslant \Big(\sum_{j=1}^m \lambda_j |f(\xi^j)|^2\Big)^{1/2} \leqslant C\epsilon^{-1}\|f\|_{L_2(\mu)}, \quad \forall f \in X_{N+1}.$$

[2]The authors are grateful to the referee for noting that the implicit constants can be made dimension-free, rather than depending on d as they did in the initial draft.

Since the constant function 1 belongs to the space X_{N+1}, this implies that

$$(\sum_{j=1}^{m} \lambda_j)^{\frac{1}{2}} \leqslant C\epsilon^{-1}.$$

It then follows from Lemma 7.6 that for any $f \in X_N$,

$$\begin{aligned}\max_{x\in\Omega} |f(x)| &\leqslant (N+1)^{\frac{1}{2}} \|f\|_{L_2(\mu)} \leqslant (N+1)^{\frac{1}{2}} \Big(\sum_{j=1}^{m} \lambda_j |f(\xi^j)|^2\Big)^{\frac{1}{2}} \\ &\leqslant (N+1)^{\frac{1}{2}} \max_{1\leqslant j\leqslant m} |f(\xi^j)| \Big(\sum_{j=1}^{m} \lambda_j\Big)^{\frac{1}{2}} \leqslant C\epsilon^{-1} N^{\frac{1}{2}} \max_{1\leqslant j\leqslant m} |f(\xi^j)|. \qquad \square\end{aligned}$$

Replacing Theorem 3.12 with Corollary 3.10 in the argument above yields the following result.

Theorem 7.7. *Given any real N-dimensional subspace X_N of $\mathcal{C}(\Omega)$ and any $\epsilon \in (0,1)$, there exists a set $\{\xi^j\}_{j=1}^m \subset \Omega$ of $m \leqslant C\epsilon^{-2}N$ points such that*

$$\max_{x\in\Omega} |f(x)| \leqslant (1+\epsilon)\sqrt{N} \max_{1\leqslant j\leqslant m} |f(\xi^j)|,$$

where $C > 1$ is an absolute constant.

We point out that Theorems 7.5 and 7.7 were proved in [31, Theorem 6.6] and [38] under the extra condition $X_N \in \mathrm{NI}_{2,\infty}(H)$ for some constant $H > 0$.

Acknowledgments

The authors sincerely thank the anonymous referees for their time and insightful comments. Their suggestions have been fully incorporated into the revised manuscript.

References

[1] M. R. Abdollahpour, M. H. Faroughi, and A. Rahimi. *PG*-frames in Banach spaces. *Methods Funct. Anal. Topology*, 13:201–210, 2007.

[2] F. Bartel, L. Kämmerer, D. Potts, and T. Ullrich. On the reconstruction of functions from values at subsampled quadrature points. *Math. Comp.*, 93(346):785–809, 2024.

[3] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 798–805. IEEE, 2008.

[4] S. N. Bernstein. Sur une classe de formules d'interpolation. *Izv. Akad. Nauk SSSR*, 9:1151–1161, 1931.

[5] S. N. Bernstein. Sur une modification de la formule d'interpolation de Lagrange. *Zapiski Khar'kovskogo Matem. Tovar.*, 5:49–57, 1932.

[6] T. Bloom, L. P. Bos, J. P. Calvi, and N. Levenberg. Polynomial interpolation and approximation in $\mathbb{C}^d$. *Ann. Polon. Math.*, 106:53–81, 2012.

[7] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162(1-2):73–141, 1989.

[8] G. Brown and F. Dai. Approximation of smooth functions on compact two-point homogeneous spaces. *J. Funct. Anal.*, 220(2):401–423, 2005.

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

[10] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(3):035020, 2008.

[11] D. Coppersmith and Th. Rivlin. The growth of polynomials bounded at equally spaced points. *SIAM J. Math. Anal.*, 23(4):970–983, 1992.

[12] F. Dai. Kolmogorov and linear widths of weighted Besov classes. *J. Math. Anal. Appl.*, 315(2):711–724, 2006.

[13] F. Dai, E. Kosov, and N. Murasko. Empirical approximation of L_p norms, 2026. arXiv:2606.00347.

[14] F. Dai, E. Kosov, and V. Temlyakov. Some improved bounds in sampling discretization of integral norms. *J. Funct. Anal.*, 285(4):109951, 2023.

[15] F. Dai and A. Prymak. Optimal polynomial meshes exist on any multivariate convex domain. *Found. Comput. Math.*, 24(3):989–1018, 2024.

[16] F. Dai, A. Prymak, A. Shadrin, V. Temlyakov, and S. Tikhonov. Entropy numbers and Marcinkiewicz-type discretization. *J. Funct. Anal.*, 281(6):109090, 2021.

[17] F. Dai, A. Prymak, A. Shadrin, V. Temlyakov, and S. Tikhonov. Sampling discretization of integral norms. *Constr. Approx.*, 54(3):455–471, 2021.

[18] F. Dai, A. Prymak, V. Temlyakov, and S. Tikhonov. Integral norm discretization and related problems. *Russian Math. Surveys*, 74(4):579–630, 2019. Translation from Uspekhi Mat. Nauk, 74(4)(448), 3–58, 2019.

[19] F. Dai and V. Temlyakov. Sampling discretization of integral norms and its application. *Proc. Steklov Inst. Math.*, 319:97–109, 2022.

[20] F. Dai and V. Temlyakov. Universal sampling discretization. *Constr. Approx.*, 58(3):589–613, 2023.

[21] F. Dai and V. Temlyakov. Random points are good for universal discretization. *J. Math. Anal. Appl.*, 529(1):127570, 2024.

[22] F. Dai and V. Temlyakov. Universal discretization and sparse recovery. *J. Approx. Theory*, 311:106199, 2025.

[23] F. Dai and V. Temlyakov. Lebesgue-type inequalities in sparse sampling recovery. *J. Math. Sci.*, 299:344–362, 2026.

[24] F. Dai and V. Temlyakov. A survey on sampling recovery, 2026. arXiv:2601.08787v1.

[25] M. Dolbeault, D. Krieg, and M. Ullrich. A sharp upper bound for sampling numbers in L_2. *Appl. Comput. Harmon. Anal.*, 63:113–134, 2023.

[26] D. E. Edmunds and H. Triebel. *Function spaces, entropy numbers, differential operators*, volume 120 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[27] H. Ehlich and K. Zeller. Schwankung von polynomen zwischen gitterpunkten. *Math. Z.*, 86(1):41–44, 1964.

[28] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, New York, 2013.

[29] T. Jahn, T. Ullrich, and F. Voigtlaender. Sampling numbers of smoothness classes via ℓ^1-minimization. *J. Complexity*, 79:101786, 2023.

[30] W. B. Johnson and G. Schechtman. Finite dimensional subspaces of L_p. In *Handbook of the Geometry of Banach Spaces*, volume 1, pages 837–870. North-Holland, Amsterdam, 2001.

[31] B. S. Kashin, S. V. Konyagin, and V. Temlyakov. Sampling discretization of the uniform norm. *Constr. Approx.*, 57(2):663–694, 2023.

[32] B. S. Kashin, E. Kosov, I. Limonova, and V. Temlyakov. Sampling discretization and related problems. *J. Complexity*, 71:101653, 2022.

[33] B. S. Kashin and V. Temlyakov. Observations on discretization of trigonometric polynomials with given spectrum. *Russian Math. Surveys*, 73(6):1128–1130, 2018. Translation from Uspekhi Mat. Nauk, 73(6), 197–198, 2018.

[34] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canad. J. Math.*, 12:363–366, 1960.

[35] C. K. Kobindarajah and D. S. Lubinsky. Marcinkiewicz–Zygmund type inequalities for all arcs of the circle. In *Advances in Constructive Approximation: Vanderbilt 2003*, Mod. Methods Math., pages 255–264. Nashboro Press, Brentwood, TN, 2004.

[36] E. Kosov. Marcinkiewicz-type discretization of L_p-norms under the Nikolskii-type inequality assumption. *J. Math. Anal. Appl.*, 504(1):125358, 2021.

[37] E. Kosov. Remarks on sampling discretization of integral norms of functions. *Proc. Steklov Inst. Math.*, 319:188–198, 2022.

[38] E. Kosov and V. Temlyakov. Sampling discretization of the uniform norm and applications. *J. Math. Anal. Appl.*, 538(2):128431, 2024.

[39] D. Krieg, K. Pozharska, M. Ullrich, and T. Ullrich. Sampling projections in the uniform norm. *J. Math. Anal. Appl.*, 553(2):129873, 2026.

[40] D. Krieg and M. Sonnleitner. Random points are optimal for the approximation of Sobolev functions. *IMA J. Numer. Anal.*, 44(3):1346–1371, 2024.

[41] D. Krieg and M. Ullrich. Function values are enough for L_2-approximation. *Found. Comput. Math.*, 21(4):1141–1151, 2021.

[42] D. Krieg and M. Ullrich. Approximation of functions: Optimal sampling and complexity. *Acta Numer.*, 35:273–457, 2026.

[43] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes.* Classics in Mathematics. Springer, Berlin, Heidelberg, 2013. Electronic reissue of the 1991 edition.

[44] D. Lewis. Finite dimensional subspaces of L_p. *Studia Math.*, 63(2):207–212, 1978.

[45] I. Limonova, Yu. Malykhin, and V. Temlyakov. One-sided discretization inequalities and sampling recovery. *Russian Math. Surveys*, 79(3):515–545, 2024. Translation from Uspekhi Mat. Nauk 79(3(477)), 149–180, 2024.

[46] I. Limonova and V. Temlyakov. On sampling discretization in L_2. *J. Math. Anal. Appl.*, 515(2):126457, 2022.

[47] D. S. Lubinsky. Marcinkiewicz–Zygmund inequalities: methods and results. In G. Milovanovic et al., editors, *Recent Progress in Inequalities*. Kluwer Academic Publishers, Dordrecht, 1998.

[48] D. S. Lubinsky. On converse Marcinkiewicz–Zygmund inequalities in L_p, $p > 1$. *Constr. Approx.*, 15(4):577–610, 1999.

[49] A. A. Lunin. Operator norms of submatrices. *Math. Notes*, 45(3):248–252, 1989. Translated from Mat. Zametki 45(3), 94–100, 1989.

[50] A. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem. *Ann. of Math. (2)*, 182(1):327–350, 2015.

[51] G. Mastroianni and V. Totik. Weighted polynomial inequalities with doubling and A_∞ weights. *Constr. Approx.*, 16(1):37–71, 2000.

[52] V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Prilozen*, 5:28–37, 1971.

[53] N. Nagel, M. Schäfer, and T. Ullrich. A new upper bound for sampling numbers. *Found. Comput. Math.*, 22(2):445–468, 2022.

[54] E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*, volume 1349 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1988.

[55] M. Rudelson. Almost orthogonal submatrices of an orthogonal matrix. *Israel J. Math.*, 111:143–155, 1999.

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

[57] G. Schechtman. Tight embedding of subspaces of L_p in ℓ_p^n for even p. *Proc. Amer. Math. Soc.*, 139(12):4419–4421, 2011.

[58] M. Talagrand. Embedding subspaces of L_1 into ℓ_1^n. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

[59] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics*. Springer, Heidelberg, 2014.

[60] V. Temlyakov. The Marcinkiewicz-type discretization theorems for the hyperbolic cross polynomials. *Jaen J. Approx.*, 9(1):37–63, 2017.

[61] V. Temlyakov. The Marcinkiewicz-type discretization theorems. *Constr. Approx.*, 48(2):337–369, 2018.

[62] V. Temlyakov. *Multivariate Approximation*, volume 32 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2018.

[63] V. Temlyakov. Universal discretization. *J. Complexity*, 47:97–109, 2018.

[64] V. Temlyakov. Sparse sampling recovery by greedy algorithms. *IMA J. Numer. Anal.*, 46(4):2107–2138, 2026.

[65] Vladimir Temlyakov. *Greedy approximation*, volume 20 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2011.

[66] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

[67] M. Ullrich. Sampling and entropy numbers in the uniform norm. *J. Complexity*, 92:101992, 2026.

[68] A. Zygmund. *Trigonometric Series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. Reprint of the 1979 edition.

Feng Dai
Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Canada
fdai@ualberta.ca

Egor D. Kosov
Centre de Recerca Matemàtica
Campus de Bellaterra
Barcelona, Spain
kosoved09@gmail.com

Vladimir N. Temlyakov
Steklov Mathematical Institute
Russian Academy of Sciences
Moscow, Russia
and
Department of Mechanics and Mathematics
Lomonosov Moscow State University
Moscow, Russia
and
Department of Mathematics
University of South Carolina
Columbia, USA
temlyakovv@gmail.com