Constructive Approximation (2023) 58:589–613
https://doi.org/10.1007/s00365-023-09644-2

CONSTRUCTIVE APPROXIMATION

Universal Sampling Discretization

F. Dai[1] · **V. Temlyakov**[2,3,4,5]

Received: 27 July 2021 / Revised: 17 March 2023 / Accepted: 20 March 2023 /
Published online: 25 April 2023

Abstract
Let X_N be an N-dimensional subspace of L_2 functions on a probability space (Ω, μ) spanned by a uniformly bounded Riesz basis Φ_N. Given an integer $1 \le v \le N$ and an exponent $1 \le p \le 2$, we obtain universal discretization for the integral norms $L_p(\Omega, \mu)$ of functions from the collection of all subspaces of X_N spanned by v elements of Φ_N with the number m of required points satisfying $m \ll v(\log N)^2(\log v)^2$. This last bound on m is much better than previously known bounds which are quadratic in v. Our proof uses a conditional theorem on universal sampling discretization, and an inequality of entropy numbers in terms of greedy approximation with respect to dictionaries.

Keywords Sampling discretization · Universality · Entropy numbers

Mathematics Subject Classification Primary 65J05; Secondary 42A05 · 65D30 · 41A63

Communicated by Ronald A. DeVore.

The first named author's research was partially supported by NSERC of Canada Discovery Grant RGPIN-2020-03909. The second named author's research wa s supported by the Russian Federation Government Grant No. 14.W03.31.0031.

✉ F. Dai
fdai@ualberta.ca

V. Temlyakov
temlyak@math.sc.edu

1 Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

2 University of South Carolina, Columbia, USA

3 Steklov Institute of Mathematics, Moscow, Russia

4 Lomonosov Moscow State University, Moscow, Russia

5 Moscow Center for Fundamental and Applied Mathematics, Moscow, Russia

1 Introduction

A standard approach to solving a continuous problem numerically – the Galerkin method – suggests looking for an approximate solution from a given finite-dimensional subspace. A typical way to measure an error of approximation is an appropriate L_p norm, $1 \le p \le \infty$. Thus, the problem of discretization of the L_p norms of functions from a given finite-dimensional subspace arises in a very natural way. Approximation by elements from a linear subspace falls in the category of *linear approximation*.

It was understood in numerical analysis and approximation theory that in many problems from signal/image processing it is beneficial to use an m-term approximant with respect to a given system of elements (dictionary) $\mathcal{D}_N := \{g_i\}_{i=1}^N$. This means that for $f \in X$ we look for an approximant of the form

$$a_m(f) := \sum_{k \in \Lambda(f)} c_k g_k \tag{1.1}$$

where $\Lambda(f) \subset [1, N]$ is a set of m indices which is determined by f. The complexity of this approximant is characterized by the cardinality $|\Lambda(f)| = m$ of $\Lambda(f)$. Approximation of this type is referred to as *nonlinear approximation* because, for a fixed m, the approximant $a_m(f)$ comes from different linear subspaces spanned by g_k, $k \in \Lambda(f)$, which depend on f. The cardinality $|\Lambda(f)|$ is a fundamental characteristic of $a_m(f)$ called *sparsity* of $a_m(f)$ with respect to $\mathcal{D}_N$. It is now well understood that we need to study nonlinear sparse approximation in order to significantly increase our ability to process (compress, denoise, etc.) large data sets. Sparse approximations of a function are not only a powerful analytic tool but they are utilized in many applications in image/signal processing and numerical computation.

Therefore, here is an important ingredient of the discretization problem, desirable in practical applications. Suppose we have a finite dictionary $\mathcal{D}_N := \{g_j\}_{j=1}^N$ of functions from $L_p(\Omega, \mu)$. Applying our strategy of sparse m-term approximation with respect to $\mathcal{D}_N$ we obtain a collection of all subspaces spanned by at most m elements of $\mathcal{D}_N$ as a possible source of approximating (representing) elements. Thus, we would like to build a discretization scheme, which works well for all such subspaces. This kind of discretization falls in the category of *universal discretization*. The paper is devoted to the problem of universal sampling discretization.

Let Ω be a nonempty set equipped with a probability measure μ. For $1 \le p \le \infty$, let $L_p(\Omega) := L_p(\Omega, \mu)$ denote the real Lebesgue space L_p defined with respect to the measure μ on Ω, and let $\|\cdot\|_p$ be the norm of $L_p(\Omega)$. By discretization of the L_p norm we understand a replacement of the measure μ by a discrete measure μ_m with support on a set $\xi = \{\xi^j\}_{j=1}^m \subset \Omega$. This means that integration with respect to measure μ is replaced by an appropriate cubature formula. Thus, integration is replaced by evaluation of a function f at a finite set ξ of points. This is why this way of discretization is called *sampling discretization*. The problem of sampling discretization is a classical problem. The first results in this direction were obtained in the 1930 s by Bernstein, by Marcinkiewicz, and by Marcinkiewicz and Zygmund for discretization of the L_p norms of the univariate trigonometric polynomials. Even though this problem is very important in applications, its systematic study has begun only recently (see the survey paper [5]). We now give explicit formulations of the sampling discretization problem (also known as the Marcinkiewicz discretization problem) and of the problem of universal discretization.

The sampling discretization problem. Let (Ω, μ) be a probability space and let $X_N \subset L_p$ be an N-dimensional subspace of $L_p(\Omega, \mu)$ with $1 \le p \le \infty$ (the index N here, usually, stands for the dimension of X_N). We shall always assume that every function in X_N is defined everywhere on Ω, and

$$f \in X_N,\ \|f\|_p = 0 \implies f = 0 \in X_N.$$

We say that X_N admits the Marcinkiewicz-type discretization with parameters $m \in \mathbb{N}$ and p and positive constants $C_1 \le C_2$ if there exists a set $\xi := \{\xi^j\}_{j=1}^m \subset \Omega$ such that for any $f \in X_N$ we have in the case of $1 \le p < \infty$,

$$C_1\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^{m}|f(\xi^j)|^p \le C_2\|f\|_p^p, \tag{1.2}$$

and in the case of $p = \infty$,

$$C_1\|f\|_\infty \le \max_{1\le j\le m}|f(\xi^j)| \le \|f\|_\infty.$$

The problem of universal discretization. Let $\mathcal{X} := \{X(n)\}_{n=1}^k$ be a collection of finite-dimensional linear subspaces $X(n)$ of the space $L_p(\Omega)$ for a given $1 \le p \le \infty$. We say that a set $\xi := \{\xi^j\}_{j=1}^m \subset \Omega$ provides *universal discretization* for the collection $\mathcal{X}$ if there are two positive constants C_i, $i = 1, 2$, such that for each $n \in \{1, \dots, k\}$ and any $f \in X(n)$ we have

$$C_1\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^{m}|f(\xi^j)|^p \le C_2\|f\|_p^p$$

in the case of $1 \le p < \infty$, and

$$C_1\|f\|_\infty \le \max_{1\le j\le m}|f(\xi^j)| \le \|f\|_\infty \tag{1.3}$$

in the case of $p = \infty$.

Note that the problem of universal discretization for the collection $\mathcal{X} := \{X(n)\}_{n=1}^k$ is the sampling discretization problem for the set $\cup_{n=1}^k X(n)$. Also, we point out that the concept of universality is well known in approximation theory. For instance, the reader can find a discussion of universal cubature formulas in [28], Section 6.8.

The problem of universal discretization for some special subspaces of the trigonometric polynomials was studied in [5, 29]. To describe the results in [5, 29] we need to introduce some necessary notations. First, for a given finite subset Q of $\mathbb{Z}^d$ we set

$$\mathcal{T}(Q) := \Big\{ f : f = \sum_{\mathbf{k}\in Q} c_{\mathbf{k}} e^{i(\mathbf{k},\mathbf{x})}, \;\; c_{\mathbf{k}} \in \mathbb{C}, \;\; \mathbf{k} \in Q \Big\}.$$

For $\mathbf{s} = (s_1, \cdots, s_d) \in \mathbb{Z}^d_+$ we define

$$R(\mathbf{s}) := \{\mathbf{k} \in \mathbb{Z}^d : |k_j| < 2^{s_j}, \quad j = 1, \dots, d\}.$$

The following result, proved in [29], solves the universal discretization problem for the collection

$$\mathcal{C}(n, d) := \{\mathcal{T}(R(\mathbf{s})) : s_1 + \cdots + s_d = n\}$$

of subspaces of trigonometric polynomials.

Theorem 1.1 [29] *For every $1 \le p \le \infty$ there exists a large enough constant $C(d, p)$, which depends only on d and p, such that for any $n \in \mathbb{N}$ there is a set $\xi := \{\xi^\nu\}_{\nu=1}^m \subset \mathbb{T}^d$, with $m \le C(d, p)2^n$ that provides universal discretization in L_p for the collection $\mathcal{C}(n, d)$.*

Second, for $n \in \mathbb{N}$ let

$$\Pi_n := [-2^{n-1} + 1, 2^{n-1} - 1]^d \cap \mathbb{Z}^d.$$

For a positive integer $v \le |\Pi_n|$ define

$$\mathcal{S}(v, n) := \{Q : \;\; Q \subset \Pi_n, \; |Q| = v\}.$$

Then it is easily seen that

$$|\mathcal{S}(v, n)| = \binom{|\Pi_n|}{v} < 2^{dnv}.$$

The following theorem provides universal discretization of L_1 and L_2 norms for the collection $\{\mathcal{T}(Q) : \;\; Q \in \mathcal{S}(v, n)\}$.

Theorem 1.2 [5, 27, Theorem 7.4] *For positive integers n and $1 \le v \le |\Pi_n|$ let*

$$M_p(n, v) := \begin{cases} v^2 n^{9/2}, & \textit{if } p = 1, \\ v^2 n, & \textit{if } p = 2. \end{cases}$$

Then there exist three positive constants $C_i(d)$, $i = 1, 2, 3$, such that for any $n, v \in \mathbb{N}$ with $v \le |\Pi_n|$, and for $p = 1$ and $p = 2$ there is a set $\xi = \{\xi^\nu\}_{\nu=1}^m \subset \mathbb{T}^d$, with

$m \le C_1(d)M_p(n,v)$ such that for any $f \in \cup_{Q\in\mathcal{S}(v,n)}\mathcal{T}(Q)$

$$C_2(d)\|f\|_p^p \le \frac{1}{m}\sum_{\nu=1}^{m}|f(\xi^\nu)|^p \le C_3(d)\|f\|_p^p.$$

Let us denote by $\mathcal{D}_N = \{g_i\}_{i=1}^N$ a system of functions from L_p. Denote the set of all v-term approximants with respect to $\mathcal{D}_N$ as

$$\Sigma_v(\mathcal{D}_N) := \left\{ f : f = \sum_{i\in G} c_i g_i, \quad \text{with } G \subset [1,N] \text{ such that} |G| = v \right\}.$$

Theorem 1.2 provides universal discretization for the collection $\{\mathcal{T}(Q) : Q \in \mathcal{S}(v,n)\}$, which is equivalent to the sampling discretization of the L_p norm of elements from the set $\Sigma_v(\mathcal{D}_N)$ with $N = |\Pi_n|$, $\mathcal{D}_N = \{e^{i(\mathbf{k},\mathbf{x})}\}_{\mathbf{k}\in\Pi_n}$. The proof of Theorem 1.2 in the case $p = 2$ is based on deep results on random matrices and in the case $p = 1$ is based on the chaining technique. We point out that in both cases $p = 2$ and $p = 1$ Theorem 1.2 provides universal discretization with the number of points growing as v^2.

On the other hand, while Theorem 1.1 provides universal discretization for the **subcollection** $\mathcal{C}(n,d)$ of $\{\mathcal{T}(Q) : Q \in \mathcal{S}(v, n+1)\}$ (rather than the whole collection $\{\mathcal{T}(Q) : Q \in \mathcal{S}(v, n+1)\}$) with $v = 2^n$ and

$$\mathcal{D} := \{e^{i(\mathbf{k},\mathbf{x})} : \mathbf{k} \in \mathbb{Z}^d, \ |k_j| < 2^n, \ 1 \le j \le d\Big\},$$

it gives a better estimate $m \le Cv$ on the number of points, which is linear in v, and applies to the full range of $1 \le p \le \infty$.

In this paper we prove the following estimate (see below for definitions and notations).

Theorem 1.3 *Let $1 \le p \le 2$. Assume that Φ_N is a uniformly bounded Riesz basis of $X_N := \mathrm{span}(\Phi_N)$ satisfying (2.8) for some constants $0 < R_1 \le R_2$. Then for a large enough constant $C = C(p, R_1, R_2)$ and any integer $1 \le v \le N$ there exist m points $\xi^1, \cdots, \xi^m \in \Omega$ with*

$$m \le Cv(\log N)^2(\log(2v))^2$$

such that for any $f \in \Sigma_v(\Phi_N)$ we have

$$\frac{1}{2}\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^{m}|f(\xi^j)|^p \le \frac{3}{2}\|f\|_p^p. \tag{1.4}$$

In particular, Theorem 1.3 gives the order of bound

$$m \ll v(\log N)^2(\log(2v))^2,$$

which is linear in v with extra logarithmic terms in N and v. This bound is much better than previously known bounds (see Theorem 1.2), which provided quadratic in v bounds. Note that even for each individual subspace from $\{\mathcal{T}(Q) : \quad Q \in \mathcal{S}(v,n)\}$ we have the lower bound $m \geq v$ for the sampling discretization.

Finally, we point out that very recent progress related to universal discretization has been made in our follow-up papers [8, 9]. More precisely, in [8] we prove that in the setting of Theorem 1.3 independent random points $\xi_1, \cdots, \xi_m \in \Omega$ that are identically distributed according to a given probabilistic measure μ provide the universal discretization (1.4) with high probability under a slightly weaker condition than the condition in Theorem 1.3 on the number of points

$$m \ll v(\log N)(\log(2v))^2(\log(2v) + \log\log N).$$

Also, in [8] we relaxed the condition on the Riesz basis Φ_N. In [9] we show how universal discretization can be applied to deduce interesting results on sparse sampling recovery. In particular, we demonstrate that a simple greedy type algorithm based on good points for universal discretization provides good recovery in the square norm.

The rest of this paper is organized as follows. Sections 2 and 3 are devoted to estimating the entropy numbers $\varepsilon_k(\Sigma_v^p(\Phi_N), L_\infty)$ of the sets

$$\Sigma_v^p(\Phi_N) := \{f \in \Sigma_v(\Phi_N) : \|f\|_p \leq 1\}, \quad 1 \leq p \leq 2,$$

in the L_∞-norm, where Φ_N is a uniformly bounded Riesz basis of $X_N := [\Phi_N] \subset L_2$ and $1 \leq v \leq N$ is an integer. Such estimates play an important role in the proof of Theorem 1.3. To be more precise, in Sect. 2 we prove under the additional condition (2.9) on the space $X_N = \operatorname{span}(\Phi_N)$ that for $p = 2$,

$$\varepsilon_k(\Sigma_v^2(\Phi_N), L_\infty) \leq C(\log N)\Big(\frac{v}{k}\Big)^{1/2}, \quad k = 1, 2, \ldots. \tag{1.5}$$

The proof of (1.5) uses a known result from Greedy approximation in smooth Banach spaces and its connection with entropy numbers. In Sect. 3, we show how the estimate (1.5) can be extended to the case $1 \leq p < 2$ under condition (2.9). This extension step is based on a general inequality for the entropy, which is given in Lemma 3.1 and appears to be of independent interest. In Sect. 4 we prove Theorem 1.3, using the estimates on entropy numbers established in the previous two sections and a conditional theorem on sampling discretization. A main step in the proof is to show that the condition (2.9) that is assumed in our estimates of entropy numbers can be dropped in sampling discretization. The conditional Theorem 2.2 used in the proof of Theorem 1.3 is given in Sect. 2 without proof. In Sect. 5, we prove a refined conditional theorem for sampling discretization of all integral norms L_p of functions from a subset $\mathcal{W} \subset L_\infty$ satisfying certain conditions, which allows us to estimate the number of points required for the sampling discretization in terms of an integral of the ε-entropy $\mathcal{H}_\varepsilon(\mathcal{W}, L_\infty)$, $\varepsilon > 0$. This is an extension of the conditional result proved in [7, 27] for the unit ball of the space $X_N \subset L_p$. In particular, it also allows us to prove a refined version of Theorem 1.3, where the constants $\frac{1}{2}$ and $\frac{3}{2}$ in (1.4) are replaced by $1-\varepsilon$ and $1+\varepsilon$

respectively for an arbitrarily given $\varepsilon \in (0,1)$. Finally, in Sect. 6 we give a few remarks on universal sampling discretization of L_p norms for $p > 2$.

Throughout this paper the letter C denotes a general positive constant depending only on the parameters indicated as arguments or subscripts. We use the notation $|A|$ to denote the cardinality of a finite set A.

2 Some General Entropy Bounds and the Case $p = 2$

It is well known that bounds of the entropy numbers of the unit ball of an N-dimensional subspace $X_N \subset L_p$

$$X_N^p := \{f \in X_N : \|f\|_p \le 1\}$$

play an important role in sampling discretization of the L_p norm of elements of X_N (see [6, 26, 27], and [7]).

Recall the definition of entropy numbers in Banach spaces. Let X be a Banach space and $B_X(g,r)$ denote the closed ball $\{f \in X : \|f-g\| \le r\}$ with center $g \in X$ and radius $r > 0$. Given a positive number ε, the covering number $N_\varepsilon(A,X)$ of a compact set $A \subset X$ is defined as

$$N_\varepsilon(A,X) := \min\left\{ n \in \mathbb{N} : \ \exists\, g^1,\ldots,g^n \in A,\ A \subset \bigcup_{j=1}^{n} B_X(g^j,\varepsilon) \right\}.$$

We denote by $\mathcal{N}_\varepsilon(A,X)$ the corresponding minimal ε-net of the set A in X; namely, $\mathcal{N}_\varepsilon(A,X)$ is a finite subset of A such that $A \subset \bigcup_{y\in\mathcal{N}_\varepsilon(A,X)} B_X(y,\varepsilon)$ and $N_\varepsilon(A,X) = |\mathcal{N}_\varepsilon(A,X)|$. The ε-entropy $\mathcal{H}_\varepsilon(A,X)$ of the compact set A in X is defined as $\log_2 N_\varepsilon(A,X)$, and the entropy numbers $\varepsilon_k(A,X)$ of the set A in X are defined as

$$\varepsilon_k(A,X) : = \inf\{\varepsilon > 0 : \mathcal{H}_\varepsilon(A,X) \le k\}, \quad k = 1,2,\ldots.$$

The following conditional result was proved in [27] for $p = 1$ and in [6] for the full range of $1 \le p < \infty$.

Theorem 2.1 [27], [6, Theorem 1.3] *Let $1 \le p < \infty$. Suppose that a subspace $X_N \subset L_p(\Omega,\mu)$ satisfies the condition*

$$\varepsilon_k(X_N^p, L_\infty) \le B(N/k)^{1/p}, \quad 1 \le k \le N, \tag{2.1}$$

where $B \ge 1$. Then for a large enough constant $C(p)$ there exist m points $\xi^1,\cdots,\xi^m \in \Omega$ with

$$m \le C(p) N B^p (\log_2(2BN))^2$$

such that for any $f \in X_N$ *we have*

$$\frac{1}{2}\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \le \frac{3}{2}\|f\|_p^p.$$

As we explained above the problem of universal discretization of the collection $\{X(n)\}_{n=1}^k$ is equivalent to the sampling discretization of the union $\cup_{n=1}^k X(n)$ of the corresponding subsets. Therefore, instead of bounds of the entropy numbers of the unit ball X_N^p we are interested in the entropy bounds of the "unit ball"

$$\Sigma_v^p(\mathcal{D}_N) := \{f \in \Sigma_v(\mathcal{D}_N) : \|f\|_p \le 1\},$$

which is the union of the corresponding unit balls.

The following version of Theorem 2.1 follows directly from its proof.

Theorem 2.2 *Let* $1 \le p < \infty$ *and* $1 \le v \le N$. *Suppose that a dictionary* $\mathcal{D}_N$ *is such that the set* $\Sigma_v^p(\mathcal{D}_N)$ *satisfies the condition*

$$\varepsilon_k(\Sigma_v^p(\mathcal{D}_N), L_\infty) \le B_1(v/k)^{1/p}\|f\|_p, \quad 1 \le k < \infty, \tag{2.2}$$

where $B_1 \ge 1$. *Assume in addition that there exists a constant* $B_2 \ge 1$ *such that*

$$\|f\|_\infty \le B_2 v^{1/p}\|f\|_p, \ \ \forall f \in \Sigma_v^p(\mathcal{D}_N). \tag{2.3}$$

Then for a large enough constant $C(p)$ *there exist m points* $\xi^1, \cdots, \xi^m \in \Omega$ *with*

$$m \le C(p)B_1^p v(\log(2B_2 v))^2$$

such that for any $f \in \Sigma_v(\mathcal{D}_N)$ *we have*

$$\frac{3}{4}\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \le \frac{5}{4}\|f\|_p^p.$$

Theorem 2.2 also follows from a more general conditional theorem that will be proved in Sect. 5 (see Corollary 5.1).

Remark 2.1 We point out that (2.2) implies

$$\|f\|_\infty \le 3B_1 v^{1/p}\|f\|_p, \ \ \forall f \in \Sigma_v^p(\mathcal{D}_N). \tag{2.4}$$

Therefore, assumption (2.3) can be dropped with B_2 replaced by $3B_1$ in the bound on m. However, in applications the constant B_2 in (2.3) may be significantly smaller than $3B_1$. For example, if $\mathcal{D}_N$ is a uniformly bounded orthonormal system with $\max_{f\in\mathcal{D}_N}\|f\|_\infty = 1$, then we can take $B_2 = 1$.

Proof of (2.4). For $\Lambda \subset [1, N] \cap \mathbb{N}$ denote $X(\Lambda) := \operatorname{span}(g_i)_{i\in\Lambda}$ and $X(\Lambda)^p := \{f \in X(\Lambda) : \|f\|_p \le 1\}$. Clearly, (2.2) implies the same bound for each $X(\Lambda)^p$ with $|\Lambda| = v$. Thus, it is sufficient to prove (2.4) for a v-dimensional subspace X_v. With a slightly worse constant $4B_1$ instead of $3B_1$ it was proved in [7, Remark 1.1]. We now show how to get a better constant. Setting $\varepsilon_1 := \epsilon_1(X_v^p, L_\infty)$, we can find two functions $f_1, f_2 \in X_v^p$ such that $X_v^p \subset B_{L_\infty}(f_1, \varepsilon_1) \cup B_{L_\infty}(f_2, \varepsilon_1)$. Since $0 \in X_v^p$, 0 is contained in one of the two balls. Without loss of generality we may assume that $0 \in B_{L_\infty}(f_1, \varepsilon_1)$ so that $\|f_1\|_\infty \le \varepsilon_1$. Since $-f_2 \in X_v^p$, we have either $-f_2 \in B_{L_\infty}(f_1, \varepsilon_1)$ or $-f_2 \in B_{L_\infty}(f_2, \varepsilon_1)$, which implies $\|f_2\|_\infty \le 2\varepsilon_1$. It then follows that $\|f\|_\infty \le 3\varepsilon_1$ for all $f \in X_v^p$. This together with (2.2) proves (2.4). □

Theorem 2.2 motivates us to estimate the characteristics $\varepsilon_k(\Sigma_v^p(\mathcal{D}_N), L_\infty)$. We now recall some known general results, which turn out to be useful for that purpose. Let $\mathcal{D}_N = \{g_j\}_{j=1}^N$ be a system of elements of cardinality $|\mathcal{D}_N| = N$ in a Banach space X. Consider the best m-term approximations of f with respect to $\mathcal{D}_N$

$$\sigma_m(f, \mathcal{D}_N)_X := \inf_{\{c_j\};\Lambda:|\Lambda|=m} \|f - \sum_{j\in\Lambda} c_j g_j\|.$$

For a set $W \subset X$ we define

$$\sigma_m(W, \mathcal{D}_N)_X := \sup_{f\in W} \sigma_m(f, \mathcal{D}_N)_X, \quad m = 1, 2, \cdots,$$

and $\sigma_0(W, \mathcal{D}_N)_X = \sup_{f\in W} \|f\|_X$. The following Theorem 2.3 was proved in [24] (see also [28], p.331, Theorem 7.4.3).

Theorem 2.3 *Let a compact $W \subset X$ be such that there exist a system $\mathcal{D}_N \subset X$ with $|\mathcal{D}_N| = N$, and a number $r > 0$ such that*

$$\sigma_m(W, \mathcal{D}_N)_X \le (m+1)^{-r}, \quad m = 0, 1, \cdots, N.$$

Then for $k \le N$

$$\varepsilon_k(W, X) \le C(r) \left(\frac{\log(2N/k)}{k}\right)^r. \tag{2.5}$$

For a given set $\mathcal{D}_N = \{g_j\}_{j=1}^N$ of elements we introduce the octahedron (generalized octahedron)

$$A_1(\mathcal{D}_N) := \left\{ f : f = \sum_{j=1}^N c_j g_j, \quad \sum_{j=1}^N |c_j| \le 1 \right\} \tag{2.6}$$

and the norm $\|\cdot\|_A$ on X_N

$$\|f\|_A := \inf \left\{ \sum_{j=1}^N |c_j| : f = \sum_{j=1}^N c_j g_j \right\}.$$

We now use a known general result for a smooth Banach space. For a Banach space X we define the modulus of smoothness

$$\rho(u) := \rho(X, u) := \sup_{\|x\|=\|y\|=1} \left(\frac{1}{2}(\|x+uy\| + \|x-uy\|) - 1\right).$$

The uniformly smooth Banach space is the one with the property

$$\lim_{u\to 0} \rho(u)/u = 0.$$

In this paper we only consider uniformly smooth Banach spaces with power type moduli of smoothness $\rho(u) \le \gamma u^s$, $1 < s \le 2$. The following bound is a corollary of greedy approximation results (see, for instance [28], p.455).

Theorem 2.4 *Let X be s-smooth: $\rho(X, u) \le \gamma u^s$, $1 < s \le 2$. Then for any normalized system $\mathcal{D}_N$ of cardinality $|\mathcal{D}_N| = N$ we have*

$$\sigma_m(A_1(\mathcal{D}_N), X) \le C(s)\gamma^{1/s} m^{1/s-1}.$$

Note that it is known that in the case $X = L_p$ we have

$$\rho(L_p, u) \le (p-1)u^2/2, \quad 2 \le p < \infty. \tag{2.7}$$

We now proceed to a special case when $X = L_p$ and $\mathcal{D}_N = \Phi_N := \{\varphi_j\}_{j=1}^N$ is a uniformly bounded Riesz basis of $X_N := [\Phi_N] := \operatorname{span}(\varphi_1, \dots, \varphi_N)$. Namely, we assume that $\|\varphi_j\|_\infty \le 1$, $1 \le j \le N$ and for any $(a_1, \cdots, a_N) \in \mathbb{R}^N$

$$R_1 \left(\sum_{j=1}^N |a_j|^2\right)^{1/2} \le \left\|\sum_{j=1}^N a_j \varphi_j\right\|_2 \le R_2 \left(\sum_{j=1}^N |a_j|^2\right)^{1/2}, \tag{2.8}$$

where $0 < R_1 \le R_2 < \infty$. Assume in addition that for any $f \in X_N$ we have

$$\|f\|_\infty \le C_0 \|f\|_{\log N}. \tag{2.9}$$

Theorem 2.5 *Assume that Φ_N is a uniformly bounded Riesz basis of $X_N := [\Phi_N]$ satisfying (2.9). Then we have*

$$\varepsilon_k(\Sigma_v^2(\Phi_N), L_\infty) \le C(R_1, C_0)(\log N)\left(\frac{v}{k}\right)^{1/2}, \quad k = 1, 2, \dots. \tag{2.10}$$

Proof First of all, for any $f = \sum_{j\in G} a_j\varphi_j$, $|G| = v$ we get

$$\|f\|_A \le \sum_{j\in G} |a_j| \le v^{1/2} \left(\sum_{j\in G} |a_j|^2\right)^{1/2} \le R_1^{-1} v^{1/2} \|f\|_2. \tag{2.11}$$

Therefore,

$$\Sigma_v^2(\Phi_N) \subset R_1^{-1} v^{1/2} \Sigma_v^A(\Phi_N),$$

where

$$R\Sigma_v^A(\Phi_N) := \{f \in \Sigma_v(\Phi_N) : \|f\|_A \le R\}.$$

By Theorem 2.4 with $s = 2$ and by (2.7) we have that for $p \in [2, \infty)$

$$\sigma_m(\Sigma_v^2(\Phi_N), L_p) \le C R_1^{-1} v^{1/2} \sqrt{p} m^{-\frac{1}{2}}, \quad m = 1, 2, \cdots, N. \tag{2.12}$$

Thus, Theorem 2.3 implies that for $p \in [2, \infty)$

$$\varepsilon_k(\Sigma_v^2(\Phi_N), L_p) \le C(R_1)(p \log(2N/k))^{1/2}(v/k)^{1/2}, \quad k = 1, 2, \ldots, N. \tag{2.13}$$

Second, by (2.9) we obtain

$$\varepsilon_k(\Sigma_v^2(\Phi_N), L_\infty) \le C_0 \varepsilon_k(\Sigma_v^2(\Phi_N), L_{\log N}). \tag{2.14}$$

Combining (2.13) and (2.14) we get

$$\varepsilon_k(\Sigma_v^2(\Phi_N), L_\infty) \le C(R_1, C_0)(\log N)(v/k)^{1/2}, \quad k = 1, 2, \ldots, N. \tag{2.15}$$

Finally, for $k > N$ we use the inequalities

$$\varepsilon_k(W, L_\infty) \le \varepsilon_N(W, L_\infty)\varepsilon_{k-N}(X_N^\infty, L_\infty)$$

and

$$\varepsilon_n(X_N^\infty, L_\infty) \le 3(2^{-n/N}), \qquad 2^{-x} \le 1/x, \quad x \ge 1,$$

to obtain (2.15) for all k. This completes the proof. □

3 A Step From $p = 2$ to $1 \le p < 2$

In this section we show how Theorem 2.5 proved in Sect. 2 for $p = 2$ can be extended to the case $1 \le p < 2$. This extension step is based on a general inequality for the entropy. For convenience, we set $\Sigma_v(\mathcal{D}_N) = X_N := [\mathcal{D}_N]$ for $v > N$.

Lemma 3.1 *For $v = 1, 2, \ldots, N$, $1 \le p < 2 < q \le \infty$, and $\theta := (\frac{1}{2} - \frac{1}{q})/(\frac{1}{p} - \frac{1}{q})$ we have for $\varepsilon > 0$*

$$\mathcal{H}_{\varepsilon}(\Sigma_v^p(\mathcal{D}_N); L_q) \le \sum_{s=0}^{\infty} \mathcal{H}_{2^{-3}a^{s-1}\varepsilon^{\theta}}(\Sigma_{2v}^2(\mathcal{D}_N); L_q) + \mathcal{H}_{\varepsilon^{\theta}}(\Sigma_{2v}^2(\mathcal{D}_N); L_q), \quad (3.1)$$

where $a = a(\theta) = 2^{\frac{\theta}{1-\theta}}$.

Proof In the case when $v = N$ Lemma 3.1 was proved in [7, Lemma 3.3]. A slight modification of the proof there works equally well for a general case of $1 \le v \le N$. For completeness, we include the proof of this lemma here. First, we note that for any $\varepsilon_1, \varepsilon_2 > 0$

$$\mathcal{H}_{\varepsilon_1\varepsilon_2}(\Sigma_v^p(\mathcal{D}_N); L_q) \le \mathcal{H}_{\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2) + \mathcal{H}_{\varepsilon_2}(\Sigma_{2v}^2(\mathcal{D}_N); L_q). \quad (3.2)$$

To see this, let $x_1, \cdots, x_{N_1} \in \Sigma_v^p(\mathcal{D}_N)$ and $y_1, \cdots, y_{N_2} \in \Sigma_{2v}^2(\mathcal{D}_N)$ be such that

$$\Sigma_v^p(\mathcal{D}_N) \subset \bigcup_{i=1}^{N_1} \left(x_i + \varepsilon_1 B_{L_2}\right) \text{ and } \Sigma_{2v}^2(\mathcal{D}_N) \subset \bigcup_{j=1}^{N_2} (y_j + \varepsilon_2 B_{L_q}),$$

where $N_1 = N_{\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N), L_2)$ and $N_2 = N_{\varepsilon_2}(\Sigma_{2v}^2(\mathcal{D}_N), L_q)$. Since $\Sigma_v(\mathcal{D}_N) + \Sigma_v(\mathcal{D}_N) \subset \Sigma_{2v}(\mathcal{D}_N)$, we have

$$\Sigma_v^p(\mathcal{D}_N) \subset \bigcup_{i=1}^{N_1} \left(x_i + \varepsilon_1 B_{L_2}\right) \cap \Sigma_v(\mathcal{D}_N) \subset \bigcup_{i=1}^{N_1} \left(x_i + \varepsilon_1 \Sigma_{2v}^2(\mathcal{D}_N)\right)$$
$$\subset \bigcup_{i=1}^{N_1} \bigcup_{j=1}^{N_2} \left(x_i + \varepsilon_1 y_j + \varepsilon_1\varepsilon_2 B_{L_q}\right).$$

Inequality (3.2) then follows.

Next, setting $\varepsilon_1 := \varepsilon^{1-\theta}$ and $\varepsilon_2 = \varepsilon^{\theta}$ in (3.2), we reduce the problem to showing that

$$\mathcal{H}_{\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2) \le \sum_{s=0}^{\infty} \mathcal{H}_{2^{-3}a^{s-1}\varepsilon^{\theta}}(\Sigma_{2v}^2(\mathcal{D}_N); L_q). \quad (3.3)$$

It will be shown that for $s = 0, 1, \ldots,$

$$\mathcal{H}_{2^s\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2) - \mathcal{H}_{2^{s+1}\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2) \le \mathcal{H}_{2^{-3}a^{s-1}\varepsilon^{\theta}}(\Sigma_{2v}^2(\mathcal{D}_N); L_q), \quad (3.4)$$

from which (3.3) will follow by taking the sum over $s = 0, 1, \ldots$

To show (3.4), for each nonnegative integer s let $\mathcal{F}_s \subset \Sigma_v^p(\mathcal{D}_N)$ be a maximal $2^s\varepsilon_1$-separated subset of $\Sigma_v^p(\mathcal{D}_N)$ in the metric L_2; that is $\|f - g\|_2 \ge 2^s\varepsilon_1$ for any two distinct functions $f, g \in \mathcal{F}_s$ and $\Sigma_v^p(\mathcal{D}_N) \subset \bigcup_{f\in\mathcal{F}_s} B_{L_2}(f, 2^s\varepsilon_1)$. Then

$$\mathcal{H}_{2^s\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2) \le \log_2|\mathcal{F}_s| \le \mathcal{H}_{2^{s-1}\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2). \quad (3.5)$$

Let $f_s \in \mathcal{F}_{s+2}$ be such that

$$\left|B_{L_2}(f_s, 2^{s+2}\varepsilon_1) \cap \mathcal{F}_s\right| = \max_{f \in \mathcal{F}_{s+2}} \left|B_{L_2}(f, 2^{s+2}\varepsilon_1) \cap \mathcal{F}_s\right|.$$

Since

$$\mathcal{F}_s = \bigcup_{f \in \mathcal{F}_{s+2}} \left(B_{L_2}(f, 2^{s+2}\varepsilon_1) \cap \mathcal{F}_s\right) \subset \Sigma_v^p(\mathcal{D}_N),$$

it follows that

$$|\mathcal{F}_s| \leq |\mathcal{F}_{s+2}| \left|B_{L_2}(f_s, 2^{s+2}\varepsilon_1) \cap \mathcal{F}_s\right|. \tag{3.6}$$

Set

$$\mathcal{A}_s := \left\{\frac{f - f_s}{2^{s+2}\varepsilon_1} : \quad f \in B_{L_2}(f_s, 2^{s+2}\varepsilon_1) \cap \mathcal{F}_s\right\} \subset \Sigma_{2v}(\mathcal{D}_N).$$

Clearly, for any $g \in \mathcal{A}_s$

$$\|g\|_2 \leq 1, \quad \|g\|_p \leq (2^{s+1}\varepsilon_1)^{-1}. \tag{3.7}$$

On the one hand, using (3.5) and (3.6), we obtain that

$$\begin{aligned} \log_2 |\mathcal{A}_s| &\geq \log_2 |\mathcal{F}_s| - \log_2 |\mathcal{F}_{s+2}| \\ &\geq \mathcal{H}_{2^s\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2) - \mathcal{H}_{2^{s+1}\varepsilon_1}(\Sigma_v^p(\mathcal{D}_N); L_2). \end{aligned} \tag{3.8}$$

On the other hand, since $\frac{1}{2} = \frac{\theta}{p} + \frac{1-\theta}{q}$, using (3.7) and the fact that $\mathcal{F}_s$ is $2^s\varepsilon_1$-separated in the L_2-metric, we have that for any two distinct $g', g \in \mathcal{A}_s$

$$2^{-2} \leq \|g' - g\|_2 \leq \|g' - g\|_p^\theta \|g - g'\|_q^{1-\theta} \leq \left(2^{s+1}\varepsilon_1\right)^{-\theta} \|g - g'\|_q^{1-\theta},$$

which implies that

$$\|g' - g\|_q \geq 2^{-2}(2^{s-1}\varepsilon_1)^{\frac{\theta}{1-\theta}} = 2^{-2}a^{s-1}\varepsilon^\theta.$$

This together with (3.7) means that $\mathcal{A}_s$ is a $2^{-2}a^{s-1}\varepsilon^\theta$-separated subset of $\Sigma_{2v}^2(\mathcal{D}_N)$ in the metric L_q. We obtain

$$\log_2 |\mathcal{A}_s| \leq \mathcal{H}_{2^{-3}a^{s-1}\varepsilon^\theta}(\Sigma_{2v}^2(\mathcal{D}_N); L_q). \tag{3.9}$$

Thus, combining (3.9) with (3.8), we prove inequality (3.4). □

Lemma 3.1 with $1 \leq p < 2$, $q = \infty$, $\theta = p/2$ and Theorem 2.5 imply the following bound for the entropy numbers.

Theorem 3.1 *Assume that Φ_N is a uniformly bounded Riesz basis of $X_N := [\Phi_N]$ satisfying (2.9). Then for $1 \le p \le 2$ we have*

$$\varepsilon_k(\Sigma_v^p(\Phi_N), L_\infty) \le C(p, R_1, C_0)(\log N)^{2/p}(v/k)^{1/p}, \quad k = 1, 2, \dots. \tag{3.10}$$

4 Proof of Theorem 1.3

Theorem 3.1 provides bounds on the entropy numbers $\varepsilon_k(\Sigma_v^p(\Phi_N), L_\infty)$ under additional assumption (2.9). Thus, a combination of Theorem 3.1 with Theorem 2.2 implies the statement of Theorem 1.3 under extra assumption (2.9). However, (2.9) is not assumed in Theorem 1.3. Below we give a proof of Theorem 1.3.

We need the following lemma proved in [7].

Lemma 4.1 [7, Lemma 4.3] *Let $1 \le p < \infty$ be a fixed number. Assume that X_N is an N-dimensional subspace of $L_\infty(\Omega)$ satisfying the following condition: For some parameter $\beta > 0$ and constant $K \ge 2$*

$$\|f\|_\infty \le (KN)^{\frac{\beta}{p}}\|f\|_p, \quad \forall f \in X_N. \tag{4.1}$$

Let $\{\xi_j\}_{j=1}^\infty$ be a sequence of independent random points distributed in accordance with μ. Then there exists a positive constant C_β depending only on β such that for any $0 < \varepsilon \le \frac{1}{2}$ and any integer

$$m \ge C_\beta K^\beta \varepsilon^{-2}(\log \frac{2}{\varepsilon})N^{\beta+1}\log N \tag{4.2}$$

the inequality

$$(1-\varepsilon)\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^m |f(\xi_j)|^p \le (1+\varepsilon)\|f\|_p^p \tag{4.3}$$

holds with probability $\ge 1 - m^{-N/\log K}$.

For a set $\Omega_m := \{x_1, \cdots, x_m\} \subset \Omega$ and a function $f : \Omega_m \to \mathbb{R}$ we define $\|f\|_{L_\infty(\Omega_m)} := \max_{1\le j\le m}|f(x_j)|$ and

$$\|f\|_{L_p(\Omega_m)} := \left(\frac{1}{m}\sum_{j=1}^m |f(x_j)|^p\right)^{\frac{1}{p}} \quad \text{for } p < \infty.$$

Now we turn to the proof of Theorem 1.3. Recall that we do not assume (2.9). First, since the Riesz basis $\Phi_N := \{\varphi_j\}_{j=1}^N$ is uniformly bounded by 1 on Ω, we have by (2.8)

$$\|f\|_\infty \le R_1^{-1}N^{\frac{1}{2}}\|f\|_2, \quad \forall f \in X_N,$$

which in turn implies that

$$\|f\|_\infty \le C(R_1)N^{\frac{1}{p}}\|f\|_p, \quad \forall f \in X_N, \quad 1 \le p \le 2.$$

Thus, by Lemma 4.1 with $\beta = 1$ there exists a discrete set $\Omega_{m_1} := \{\xi^1, \cdots, \xi^{m_1}\} \subset \Omega$ with

$$C^{-1}N^2 \log N \le m_1 \le CN^2 \log N$$

such that for all $f \in X_N$

$$\frac{4}{5}\|f\|_p^p \le \|f\|_{L_p(\Omega_{m_1})}^p \le \frac{6}{5}\|f\|_p^p \ \text{ and } \ \frac{4}{5}\|f\|_2^2 \le \|f\|_{L_2(\Omega_{m_1})}^2 \le \frac{6}{5}\|f\|_2^2, \tag{4.4}$$

where $C > 1$ is an absolute constant.

Second, we consider the discrete norm $\|\cdot\|_{L_p(\Omega_{m_1})}$ instead of the norm $\|\cdot\|_{L_p(\Omega)}$. By (4.4) Φ_N is a uniformly bounded Riesz basis of the space $(X_N, \|\cdot\|_{L_2(\Omega_{m_1})})$ and moreover

$$\|f\|_{L_\infty(\Omega_{m_1})} \le C(p, R_1, R_2)v^{1/p}\|f\|_{L_p(\Omega_{m_1})}, \quad \forall f \in \Sigma_v(\Phi_N).$$

Since $\log m_1 \sim \log N$, by the regular Nikolskii inequality for the norms $\ell_p^{m_1}$, $1 \le p \le \infty$, we also have

$$\|f\|_{L_\infty(\Omega_{m_1})} \le C\|f\|_{L_{\log N}(\Omega_{m_1})}, \quad \forall f \in X_N,$$

where $C > 1$ is an absolute constant. Thus, by Theorem 2.2 and Theorem 3.1 applied to the discrete norm $\|\cdot\|_{L_p(\Omega_{m_1})}$ we can find a subset $\Omega_m \subset \Omega_{m_1}$ with

$$m = |\Omega_m| \le C(p, R_1, R_2)v(\log N)^2(\log(2v))^2$$

such that for any $f \in \Sigma_v(\Phi_N)$

$$\frac{3}{4}\|f\|_{L_p(\Omega_{m_1})}^p \le \|f\|_{L_p(\Omega_m)}^p \le \frac{5}{4}\|f\|_{L_p,(\Omega_{m_1})}^p. \tag{4.5}$$

Combining (4.4) with (4.5), we obtain the stated result of Theorem 1.3.

5 A Refined Version of the Conditional Theorem

Let us first recall some notations. Let (Ω, μ) be a probability space. For $1 \le p \le \infty$ denote by $L_p(\Omega)$ the usual Lebesgue space L_p defined with respect to the measure μ on Ω and by $\|\cdot\|_p$ the norm of $L_p(\Omega)$. We also set

$$B_{L_p} := \{f \in L_p(\Omega) : \ \|f\|_p \le 1\}, \quad 1 \le p \le \infty.$$

In this section we prove a refined version of the conditional Theorem 2.2 for sampling discretization of all integral norms L_p of functions from a more general subset $\mathcal{W} \subset L_\infty$, which allows us to estimate the number of points needed for the sampling discretization in terms of an integral of the ε-entropy $\mathcal{H}_\varepsilon(\mathcal{W}, L_\infty)$, $\varepsilon > 0$.

Theorem 5.1 *Let $1 \le p < \infty$ and let $\mathcal{W}$ be a set of uniformly bounded functions on Ω with*

$$1 \le R := \sup_{f \in \mathcal{W}} \sup_{x \in \Omega} |f(x)| < \infty.$$

Assume that $\mathcal{H}_t(\mathcal{W}, L_\infty) < \infty$ for every $t > 0$, and

$$(\lambda \cdot \mathcal{W}) \cap B_{L_p} \subset \mathcal{W} \subset B_{L_p}, \quad \forall \lambda > 0. \tag{5.1}$$

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

$$m \ge C_p \varepsilon^{-5} \left(\int_{10^{-1}\varepsilon^{1/p}}^{R} u^{\frac{p}{2}-1} \left(\int_u^R \frac{\mathcal{H}_{c_p \varepsilon t}(\mathcal{W}, L_\infty)}{t} \, dt \right)^{\frac{1}{2}} du \right)^2, \tag{5.2}$$

there exist m points $x_1, \cdots, x_m \in \Omega$ such that for all $f \in \mathcal{W}$,

$$(1 - \varepsilon)\|f\|_p^p \le \frac{1}{m} \sum_{j=1}^{m} |f(x_j)|^p \le (1 + \varepsilon)\|f\|_p^p. \tag{5.3}$$

In particular, Theorem 5.1 allows us to prove refined versions of Theorem 2.2 and Theorem 1.3, where the constants in the Marcinkiewicz type discretization are replaced by $1 - \varepsilon$ and $1 + \varepsilon$ for an arbitrarily given $\varepsilon \in (0, 1)$.

First, we have the following refined version of Theorem 2.2.

Corollary 5.1 *Let $1 \le p < \infty$ and $1 \le v \le N$. Suppose that a dictionary $\mathcal{D}_N$ is such that*

$$\varepsilon_k(\Sigma_v^p(\mathcal{D}_N), L_\infty) \le B_1 (v/k)^{1/p}, \quad k = 1, 2, \cdots, \tag{5.4}$$

where $B_1 \ge 1$. Assume in addition that there exists a constant $B_2 \ge 1$ such that

$$\|f\|_\infty \le B_2 v^{1/p} \|f\|_p, \quad \forall f \in \Sigma_v^p(\mathcal{D}_N). \tag{5.5}$$

Then for a large enough constant $C(p)$ and any $\varepsilon \in (0, 1)$ there exist m points $\xi^1, \cdots, \xi^m \in \Omega$ with

$$m \le C(p) \varepsilon^{-5-p} v B_1^p (\log(B_2 v/\varepsilon))^2$$

such that for any $f \in \Sigma_v(\mathcal{D}_N)$

$$(1-\varepsilon)\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \le (1+\varepsilon)\|f\|_p^p.$$

Proof of Corollary 5.1 We apply Theorem 5.1 to $\mathcal{W} := \Sigma_v^p(\mathcal{D}_N)$ and $R = B_2 v^{1/p}$. It is clear that $\mathcal{W}$ satisfies (5.1). Furthermore, (5.4) implies that (see by [7, Lemma 2.1])

$$\mathcal{H}_t(\mathcal{W}, L_\infty) \le C(p)v \cdot (B_1/t)^p, \quad t > 0. \tag{5.6}$$

Finally, a straightforward calculation using (5.6) then shows that

$$\varepsilon^{-5}\left(\int_{10^{-1}\varepsilon^{1/p}}^{B_2 v^{1/p}} u^{\frac{p}{2}-1}\left(\int_u^{B_2 v^{1/p}} \frac{\mathcal{H}_{c_p\varepsilon t}(\mathcal{W}, L_\infty)}{t}\,dt\right)^{\frac{1}{2}} du\right)^2$$
$$\le C(p)\varepsilon^{-5-p}B_1^p v(\log(B_2 v/\varepsilon))^2.$$

Corollary 5.1 then follows from Theorem 5.1. □

Using Corollary 5.1 and following the proof in Sect. 4, we can also obtain the ε-version of Theorem 1.3.

Corollary 5.2 *Let* Φ_N *be a uniformly bounded Riesz basis of* $X_N := \text{span}(\Phi_N) \subset L_2(\Omega)$ *satisfying (2.8) for some constants* $0 < R_1 \le R_2$. *Let* $1 \le p \le 2$ *and let* $1 \le v \le N$ *be an integer. Then for a large enough constant* $C = C(p, R_1, R_2)$ *and any* $\varepsilon \in (0, 1)$ *there exist m points* $\xi^1, \cdots, \xi^m \in \Omega$ *with*

$$m \le C\varepsilon^{-p-5}v(\log N)^2(\log(2v\varepsilon^{-1}))^2$$

such that for any $f \in \Sigma_v(\Phi_N)$ *we have*

$$(1-\varepsilon)\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \le (1+\varepsilon)\|f\|_p^p.$$

The rest of this section is devoted to the proof of Theorem 5.1, which is close to the proof of Theorem 1.3 of [6]. We need the following lemma:

Lemma 5.1 [6, Lemma 2.4] *Let* $\{\mathcal{F}_j\}_{j\in G}$ *be a collection of finite sets of bounded functions from* $L_1(\Omega, \mu)$. *Assume that for each* $j \in G$ *and all* $f \in \mathcal{F}_j$ *we have*

$$\|f\|_1 \le 1, \quad \|f\|_\infty := \sup_{x\in\Omega}|f(x)| \le M_j.$$

Suppose that positive numbers $\eta_j \in (0, 1)$ *and a natural number* m *satisfy the condition*

$$2 \sum_{j \in G} |\mathcal{F}_j| \exp\left(-\frac{m\eta_j^2}{8M_j}\right) < 1.$$

Then there exists a set $\xi = \{\xi^\nu\}_{\nu=1}^m \subset \Omega$ *such that for each* $j \in G$ *and for all* $f \in \mathcal{F}_j$ *we have*

$$\left| \|f\|_1 - \frac{1}{m} \sum_{\nu=1}^m |f(\xi^\nu)| \right| \le \eta_j.$$

Proof of Theorem 5.1 Let

$$\mathcal{W}_1 := \{f/\|f\|_p : \quad f \in \mathcal{W} \setminus \{0\}\}.$$

Clearly, $\mathcal{W}_1 \subset \mathcal{W}$, and it suffices to prove (5.3) for all $f \in \mathcal{W}_1$. Let $c^* = c_p^* \in (0, \frac{1}{2})$ be a sufficiently small constant depending only on p. Let $a := c^*\varepsilon$. Let J, j_0 be two integers such that $j_0 < 0 \le J$,

$$(1+a)^{J-1} \le R < (1+a)^J \text{ and } (1+a)^{j_0 p} \le \frac{1}{5}\varepsilon \le (1+a)^{(j_0+1)p}. \tag{5.7}$$

For $j \in \mathbb{Z}$, let

$$\mathcal{A}_j := \mathcal{N}_{2a(1+a)^j}(\mathcal{W}_1, L_\infty) \subset \mathcal{W}_1$$

denote the minimal $2a(1+a)^j$-net of $\mathcal{W}_1$ in the norm of L_∞. For $j \in \mathbb{Z}$ and $f \in \mathcal{W}_1$ we define $A_j(f)$ to be the function in $\mathcal{A}_j$ that is closest to f in the L_∞ norm. Thus, $\|A_j(f) - f\|_\infty \le 2a(1+a)^j$ for all $f \in \mathcal{W}_1$ and $j \in \mathbb{Z}$.

Next, for $f \in \mathcal{W}_1$ and $j > j_0$ define

$$U_j(f) := \{\mathbf{x} \in \Omega : |A_j(f)(\mathbf{x})| \ge (1+a)^{j-1}\},$$

and

$$D_j(f) := U_j(f) \setminus \bigcup_{k \ge j+1} U_k(f).$$

We also set

$$D_{j_0}(f) := \Omega \setminus \bigcup_{k > j_0} U_k(f).$$

Note that by (5.7) $U_j(f) = \emptyset$ for $j > J$. Thus, $\{D_j(f) : \quad j = j_0, \cdots, J\}$ forms a partition of the domain Ω. Define

$$h(f, \mathbf{x}) := \sum_{j=j_0+1}^{J} (1+a)^j \chi_{D_j(f)}(\mathbf{x}), \tag{5.8}$$

where $\chi_E(\mathbf{x})$ is a characteristic function of a set E.

For $\mathbf{x} \in D_{j_0}(f)$ we have

$$\begin{aligned}|f(\mathbf{x})| &\le |A_{j_0+1} f(\mathbf{x})| + 2a(1+a)^{j_0+1} \le (1+a)^{j_0} + 2a(1+a)^{j_0+1}\\ &\le (1+a)^{j_0}(1+4a),\end{aligned}$$

which in turn implies that

$$|f(\mathbf{x})|^p \le (1+a)^{j_0 p}(1+4a)^p \le (1+a)^{j_0 p}(1+C_p a) \le \frac{\varepsilon}{10}(1+C_p a).$$

On the other hand, for $\mathbf{x} \in D_j(f)$ and $j_0 < j \le J$ we have

$$\begin{aligned}|f(\mathbf{x})| &\ge |A_j f(\mathbf{x})| - 2a(1+a)^j \ge (1+a)^j(1-3a) \quad \text{and}\\ |f(\mathbf{x})| &\le |A_{j+1} f(\mathbf{x})| + 2a(1+a)^{j+1} \le (1+a)^j(1+3a),\end{aligned}$$

which implies

$$(1+3a)^{-p}|f(\mathbf{x})|^p \le |h(f, \mathbf{x})|^p \le (1-3a)^{-p}|f(\mathbf{x})|^p.$$

Therefore, choosing $c^* = c_p^*$ small enough, we have

$$\left(1 - \frac{\varepsilon}{8}\right)|f(\mathbf{x})|^p \le |h(f, \mathbf{x})|^p \le \left(1 + \frac{\varepsilon}{8}\right)|f(\mathbf{x})|^p, \quad \forall \mathbf{x} \in \bigcup_{j_0 < j \le J} D_j(f), \tag{5.9}$$

and

$$|f(\mathbf{x})|^p \le \frac{\varepsilon}{8}, \quad \forall \mathbf{x} \in D_{j_0}(f).$$

In particular, this implies that for any probability measure ν on Ω and any $f \in \mathcal{W}_p'$

$$\left| \|h(f)\|_{L_p(\nu)}^p - \|f\|_{L_p(\nu)}^p \right| \le \frac{\varepsilon}{8}\|f\|_{L_p(\nu)}^p + \frac{\varepsilon}{8}. \tag{5.10}$$

For $j_0 + 1 \le j \le J$ let

$$\mathcal{F}_j^p := \left\{(1+a)^{pj}\chi_{D_j(f)} : \ f \in \mathcal{W}_1\right\}.$$

Our aim is to find m points $\xi^1, \cdots, \xi^m \in \Omega$ for each m satisfying (5.2) so that the following inequality holds for all $f \in \mathcal{F}_j^p$ and $j_0 < j \le J$:

$$\Big|\frac{1}{m}\sum_{k=1}^m f(\xi^k) - \int_\Omega f(x)\,d\mu(x)\Big| \le \varepsilon_j, \tag{5.11}$$

where $\{\varepsilon_j\}_{j=j_0+1}^J \subset (0,1)$ satisfies $\sum_{j=j_0+1}^J \varepsilon_j \le \varepsilon/4$. Once (5.11) is proved, we obtain by (5.8) that

$$\Big|\frac{1}{m}\sum_{j=1}^m |h(f,\xi^j)|^p - \|h(f)\|_p^p\Big| \le \frac{\varepsilon}{4}, \tag{5.12}$$

which, applying (5.10), will prove the desired inequality (5.3).

To see this, we apply Lemma 5.1 for the collection of the above sets $\mathcal{F}_j^p$ and notice that for $j_0 < j \le J$

$$\|(1+a)^{pj}\chi_{D_j(f)}\|_1 \le \|h(f)\|_p^p \le \|f\|_p^p + \frac{\varepsilon}{4} \le 2$$

and

$$\|(1+a)^{pj}\chi_{D_j(f)}\|_\infty \le (1+a)^{pj} =: M_j.$$

Thus, by Lemma 5.1 it suffices to show that for each integer m satisfying (5.2) one can find a sequence $\{\varepsilon_j\}_{j_0<j\le J} \subset (0,1)$ such that

$$\sum_{j_0<j\le J} \varepsilon_j \le \frac{\varepsilon}{4} \tag{5.13}$$

$$\sum_{j=j_0+1}^J |\mathcal{F}_j^p| \exp\Big(-\frac{m\varepsilon_j^2}{8M_j}\Big) < \frac{1}{2}. \tag{5.14}$$

To this end we need to estimate the cardinalities of the sets $\mathcal{F}_j^p$. By definition, for each $j_0 < j \le J$, the set $D_j(f)$ is uniquely determined by the functions $A_k(f) \in \mathcal{A}_k$, $j \le k \le J$. As a result, we have

$$|\mathcal{F}_j^p| \le |\mathcal{A}_j| \times \cdots \times |\mathcal{A}_J| =: L_j,$$

and

$$\log L_j \le \sum_{k=j}^J \log|\mathcal{A}_k| \le \sum_{k=j}^J \mathcal{H}_{a(1+a)^k}(\mathcal{W}, L_\infty)$$
$$\le \frac{1}{\log(1+a)} \sum_{k=j}^J \int_{a(1+a)^{k-1}}^{a(1+a)^k} \mathcal{H}_t(\mathcal{W}, L_\infty)\frac{dt}{t}$$

$$\leq C\varepsilon^{-1}\int_{(1+a)^{j-1}}^{R}\mathcal{H}_{at}(\mathcal{W},L_\infty)\frac{dt}{t}. \tag{5.15}$$

For each $j_0 < j \leq J$ we choose $\varepsilon_j > 0$ so that

$$\log(\lambda L_j) = \frac{m\varepsilon_j^2}{16M_j}, \quad \text{that is } \ \varepsilon_j := 4\sqrt{M_j}\sqrt{\log(\lambda L_j)}m^{-\frac{1}{2}}, \tag{5.16}$$

where $\lambda > 1$ is a large absolute constant to be specified later. Then

$$\begin{aligned}\sum_{j=j_0+1}^{J}\varepsilon_j &= 4m^{-1/2}\sum_{j=j_0+1}^{J}(M_j\log(\lambda L_j))^{\frac{1}{2}}\\ &\leq 4m^{-1/2}\sqrt{\log\lambda}\sum_{j=j_0+1}^{J}(M_j\log(L_j))^{\frac{1}{2}}\end{aligned}$$

and hence (5.13) is ensured once

$$m \geq \varepsilon^{-2}\Big(\sqrt{\log\lambda}\sum_{j=j_0+1}^{J}(M_j\log(L_j))^{\frac{1}{2}}\Big)^2. \tag{5.17}$$

However, using (5.15), we have

$$\begin{aligned}\sum_{j=j_0+1}^{J}(M_j\log L_j)^{\frac{1}{2}} &\leq C\varepsilon^{-\frac{1}{2}}\sum_{j=j_0+1}^{J}(1+a)^{pj/2}\Big(\int_{(1+a)^{j-1}}^{R}\mathcal{H}_{at}(\mathcal{W},L_\infty)\frac{dt}{t}\Big)^{\frac{1}{2}}\\ &\leq C_p\varepsilon^{-\frac{3}{2}}\sum_{j=j_0+1}^{J}\int_{(1+a)^{j-2}}^{(1+a)^{j-1}}u^{\frac{p}{2}-1}\Big(\int_u^R\mathcal{H}_{at}(\mathcal{W},L_\infty)\frac{dt}{t}\Big)^{\frac{1}{2}}du\\ &\leq C_p\varepsilon^{-\frac{3}{2}}\int_{10^{-1}\varepsilon^{1/p}}^{R}u^{\frac{p}{2}-1}\Big(\int_u^R\mathcal{H}_{c_p\varepsilon t}(\mathcal{W},L_\infty)\frac{dt}{t}\Big)^{\frac{1}{2}}du.\end{aligned}$$

This combined with (5.17) implies that (5.13) is ensured by (5.2).

Finally, we prove (5.14). Indeed, using (5.16), we have

$$\begin{aligned}&\sum_{j=j_0+1}^{J}|\mathcal{F}_j^p|\exp\Big(-\frac{m\varepsilon_j^2}{8M_j}\Big) \leq \lambda\sum_{j=j_0+1}^{J}L_j\exp\Big(-\frac{m\varepsilon_j^2}{8M_j}\Big)\\ &= \sum_{j=j_0+1}^{J}\exp\Big(\log(\lambda L_j)-\frac{m\varepsilon_j^2}{8M_j}\Big) = \sum_{j=j_0+1}^{J}\exp\Big(-\log(\lambda L_j)\Big) = \frac{1}{\lambda}\sum_{j=j_0+1}^{J}\frac{1}{L_j}\\ &\leq \frac{1}{\lambda}\sum_{j=j_0+1}^{J}\frac{1}{N_{a(1+a)^j}(\mathcal{W},L_\infty)},\end{aligned}$$

where the last step uses the fact that

$$L_j \geq |\mathcal{A}_j| = N_{a(1+a)^j}(\mathcal{W}, L_\infty).$$

We claim that

$$\mathcal{H}_t(\mathcal{W}, L_\infty) \geq \log \frac{R}{4t}, \quad \forall 0 < t < R. \tag{5.18}$$

To see this, let $f^* \in \mathcal{W}$ be such that $\|f^*\|_\infty = R$. Let $k = [\frac{R}{t}]$. Define $f_j = \frac{2jt}{R} f^*$ for $0 \leq j \leq k/2$. Then $\{f_j\}_{0 \leq j \leq \frac{k}{2}} \subset \mathcal{W}$ is $2t$-separated in L_∞-norm. It follows that

$$N_t(\mathcal{W}, L_\infty) \geq \frac{k}{2} \geq \frac{1}{4}\frac{R}{t},$$

which shows (5.18).

Now using (5.18), we obtain

$$\frac{1}{\lambda} \sum_{j=j_0+1}^{J} \frac{1}{N_{a(1+a)^j}(\mathcal{W}, L_\infty)} \leq C R^{-1} \lambda^{-1} \sum_{j=j_0+1}^{J} (1+a)^j \leq C\lambda^{-1} < 1,$$

provided that $\lambda > 1$ is large enough. This proves (5.14). □

6 Concluding Remarks on Sampling Discretization of L_p norms for $2 < p < \infty$

In this section, we give a few remarks on sampling discretization of L_p norms for $2 < p < \infty$.

1. The following Nikolskii type inequality plays an important role in the proof of Theorem 1.3:

$$\|f\|_\infty \leq C v^{\frac{1}{p}} \|f\|_p, \quad \forall f \in \Sigma_v(\Phi_N), \tag{6.1}$$

where the constant C is independent of f, v and N. This inequality holds for $1 \leq p \leq 2$ whenever Φ_N is a uniformly bounded Riesz basis of X_N. However, this is no longer true for $p > 2$. For example, take $N = 2^v$ and consider the system

$$\Phi_N = \{e^{2\pi i j x}\}_{j=1}^N$$

on the interval $[0, 1]$ equipped with the usual Lebesgue measure. By the Littlewood-Paley inequality we have that for $f(x) = \sum_{j=1}^{v} e^{2\pi \mathbf{i} 2^j x} \in \Sigma_v(\Phi_N)$ and $2 < p < \infty$,

$$\|f\|_\infty = v > C v^{\frac{1}{p}} \|f\|_p \asymp v^{\frac{1}{2}+\frac{1}{p}}.$$

2. Let Φ_N be a uniformly bounded Riesz basis of $X_N \subset L_2$ satisfying (2.9). By monotonicity of the L_p norms, we have that for any integer $1 \le v \le N$,

$$\Sigma_v^p(\Phi_N) \subset \Sigma_v^2(\Phi_N), \quad p > 2,$$

which in particular implies that

$$\sup_{f \in \Sigma_v^p(\Phi_N)} \|f\|_\infty \le \sup_{f \in \Sigma_v^2(\Phi_N)} \|f\|_\infty \le C v^{1/2}.$$

Moreover, using Theorem 2.5, we have that for $p > 2$ and all integer $k \ge 1$,

$$\varepsilon_k(\Sigma_v^p(\Phi_N), L_\infty) \le \varepsilon_k(\Sigma_v^2(\Phi_N), L_\infty) \le C \cdot (\log N)\Big(\frac{v}{k}\Big)^{1/2}, \tag{6.2}$$

which also yields

$$\mathcal{H}_t(\Sigma_v^p(\Phi_N), L_\infty) \le C(p)v \cdot \Big(\frac{\log N}{t}\Big)^2, \quad \forall t > 0. \tag{6.3}$$

On the other hand, a straightforward calculation shows that for any $\varepsilon \in (0, 1)$ and $p > 2$,

$$\varepsilon^{-5}\left(\int_{10^{-1}\varepsilon^{1/p}}^{Cv^{1/2}} u^{\frac{p}{2}-1}\Big(\int_u^{Cv^{1/2}} \frac{\mathcal{H}_{c_p\varepsilon t}(\Sigma_v^p(\Phi_N), L_\infty)}{t}\,dt\Big)^{\frac{1}{2}} du\right)^2 \le C(p)\varepsilon^{-7} v^{p/2} (\log N)^2.$$

Thus, an application of Theorem 5.1 leads to

Theorem 6.1 *Assume that Φ_N is a uniformly bounded Riesz basis of $X_N := \mathrm{span}(\Phi_N)$ satisfying (2.8) for some constants $0 < R_1 \le R_2$. Let $2 < p < \infty$ and let $1 \le v \le N$ be an integer. Then for a large enough constant $C = C(p, R_1, R_2)$ and any $\varepsilon \in (0, 1)$ there exist m points $\xi^1, \cdots, \xi^m \in \Omega$ with*

$$m \le C\varepsilon^{-7} v^{p/2} (\log N)^2$$

such that for any $f \in \Sigma_v(\Phi_N)$

$$(1 - \varepsilon)\|f\|_p^p \le \frac{1}{m}\sum_{j=1}^m |f(\xi^j)|^p \le (1 + \varepsilon)\|f\|_p^p.$$

Acknowledgements The authors would like to thank the referees for careful reading of the paper and for helpful suggestions and comments.

References

1. Batson, J., Spielman, D.A., Srivastava, N.: Twice-Ramanujan Sparsifiers. SIAM J. Comput. **41**, 1704–1721 (2012)
2. Belinskii, E.S.: Decomposition theorems and approximation by a "floating" system of exponentials. Trans. Amer. Math. Soc. **350**(1), 43–53 (1998)
3. Bourgain, J., Lindenstrauss, J., Milman, V.: Approximation of zonoids by zonotopes. Acta Math. **162**, 73–141 (1989)
4. Carl, B.: Entropy numbers, s-numbers, and eigenvalue problems. J. Funct. Anal. **41**, 290–306 (1981)
5. Dai, F., Prymak, A., Temlyakov, V.N., Tikhonov, S.: Integral norm discretization and related problems. Russ. Math. Surv. **74**, 579–630 (2019)
6. Dai, F., Prymak, A., Shadrin, A., Temlyakov, V.N., Tikhonov, S.: Sampling discretization of integral norms. Constr. Approx. **54**(3), 455–471 (2021)
7. Dai, F., Prymak, A., Shadrin, A., Temlyakov, V.N., Tikhonov, S.: Entropy numbers and Marcinkiewicz-type discretization. J. Funct. Anal. **281**(6), 109090 (2021)
8. Dai, F., Temlyakov, V.: Random points are good for universal discretization; arXiv:2301.12536
9. Dai, F., Temlyakov, V.: Universal discretization and sparse sampling recovery; arXiv:2301.05962
10. DeVore, R.A., Temlyakov, V.N.: Nonlinear approximation by trigonometric sums. J. Fourier Anal. Appl. **2**(1), 29–48 (1995)
11. Dũng, Ding, Temlyakov, V.N., Ullrich, T.: Hyperbolic Cross Approximation, Advanced Courses in Mathematics CRM Barcelona, Birkhäuser, (2018); arXiv:1601.03978v2
12. Hinrichs, A., Prochno, J., Vybiral, J.: Entropy numbers of embeddings of Schatten classes. J. Functional Analysis **273**, 3241–3261 (2017); arXiv:1612.08105v1
13. Johnson, W.B., Schechtman, G.: Finite dimensional subspaces of L_p, Handbook of the geometry of Banach spaces, Vol. 1, 837–870, North-Holland, Amsterdam (2001)
14. Kosov, E.: Marcinkiewicz-type discretization of L_p-norms under the Nikolskii-type inequality assumption. J. Math. Anal. Appl. **504**(1), 125358 (2021)
15. Marcus, A., Spielman, D.A., Srivastava, N.: Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. Annal. Math. **182**, 327–350 (2015)
16. Nitzan, S., Olevskii, A., Ulanovskii, A.: Exponential frames on unbounded sets. Proc. Am. Math. Soc. **144**, 109–118 (2016)
17. Pajor, A., Tomczak-Yaegermann, N.: Subspaces of small codimension of finite-dimensional Banach spaces. Proc. Am. Math. Soc. **97**, 637–642 (1986)
18. Rudelson, M.: Almost orthogonal submatrices of an orthogonal matrix. Israel J. Math. **111**, 143–155 (1999)
19. Schechtman, G.: Tight embedding of subspaces of L_p in ℓ_p^n for even p. Proc. Am. Math. Soc. **139**(12), 4419–4421 (2011)
20. Schechtman, G.: More on embedding subspaces of L_p in ℓ_r^n. Compos. Math. **61**(2), 159–169 (1987)
21. Schütt, C.: Entropy numbers of diagonal operators between symmetric Banach spaces. J. Approx. Theor. **40**, 121–128 (1984)
22. Sudakov, V.N.: Gaussian random processes and measures of solid angles in Hilbert spaces. Sov. Math. Dokl. **12**, 412–415 (1971)
23. Temlyakov, V.N.: Greedy Approximation. Cambridge University Press, Cambridge (2011)
24. Temlyakov, V.N.: An inequality for the entropy numbers and its application. J. Approx. Theor. **173**, 110–121 (2013)
25. Temlyakov, V.N.: On the entropy numbers of the mixed smoothness function classes. J. Approx. Theor. **207**, 26–56 (2017)
26. Temlyakov, V.N.: The Marcinkewiecz-type discretization theorems for the hyperbolic cross polynomials. Jaen Journal on Approximation **9**(1), 37–63 (2017); arXiv: 1702.01617v2
27. Temlyakov, V.N.: The Marcinkiewicz-type discretization theorems. Constr. Approx. **48**, 337–369 (2018)
28. Temlyakov, V.N.: Multivariate Approximation. Cambridge University Press, Cambridge (2018)
29. Temlyakov, V.N.: Universal discretization. J. Complex. **47**, 97–109 (2018)
30. Temlyakov, V.N.: A remark on entropy numbers. Studia Math. **263**(2), 199–208 (2022); arXiv:2008.13030
31. Zygmund, A.: Trigonometric Series. Cambridge University Press, Cambridge (1959)