Applied Numerical Mathematics 200 (2024) 209–235

Contents lists available at ScienceDirect

Applied Numerical Mathematics

journal homepage: www.elsevier.com/locate/apnum

Encoding of data sets and algorithms

Katarina Doctor [a], Tong Mao [b,1], Hrushikesh Mhaskar [b,*,2]

[a] Navy Center for Applied Research in AI, Information Techchology Division, U.S. Naval Research Laboratory, Washington, DC 20375, United States

[b] Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711, United States

A R T I C L E I N F O

Article history:
Received 15 February 2023
Received in revised form 24 June 2023
Accepted 13 July 2023
Available online 18 July 2023

Keywords:
Metric entropy
Covering number
Analytic functions
Entire functions
Entropy of class of functionals

A B S T R A C T

In many high-impact applications, it is important to ensure the quality of the output of a machine learning algorithm as well as its reliability in comparison to the complexity of the algorithm used. In this paper, we have initiated a mathematically rigorous theory to decide which models (algorithms applied on data sets) are close to each other in terms of certain metrics, such as performance and the complexity level of the algorithm. This involves creating a grid on the hypothetical spaces of data sets and algorithms so as to identify a finite set of probability distributions from which the data sets are sampled and a finite set of algorithms. A given threshold metric acting on this grid will express the nearness (or statistical distance) of each algorithm and data set of interest to any given application. A technically difficult part of this project is to estimate the so-called metric entropy of a compact subset of functions of **infinitely many variables** that arise in the definition of these spaces.

1. Introduction

In many high-impact applications of machine learning, data is limited and training is challenging. For these applications, it is desirable to have predictions with the highest assurances from the available data while minimizing uncertainty. In particular, it is important to ensure the quality of the output of a machine learning algorithm as well as its reliability compared to the complexity of the algorithm used. The objective of this work is to develop a systematic and mathematically rigorous approach to decide what is the complexity level of the algorithm that is sufficient in the task domain to produce the desired performance, reliability, and uncertainty.

An attractive idea in this context is that of Rashomon curves [5,1]. The question is the following: if one finds that an algorithm with a certain complexity level works well on a task on a data set, are there likely to be simpler algorithms that will also work within a certain tolerance of this algorithm? More generally, which class of algorithms can be expected to behave similarly on which kind of data sets? Unfortunately, there seems to be no mathematically precise formulation of this problem. Our purpose in this paper is to initiate such a rigorous study.

* Corresponding author.
E-mail addresses: katarina.doctor@nrl.navy.mil (K. Doctor), tong.mao@cgu.edu (T. Mao), hrushikesh.mhaskar@cgu.edu (H. Mhaskar).

1 The research of this author was funded by NSF DMS grant 2012355.

2 The research of HNM was supported in part by ARO grant W911NF2110218, NSF DMS grant 2012355, ONR grant N00014-23-1-2394, and a Faculty Visiting Fellowship program at ONR.

https://doi.org/10.1016/j.apnum.2023.07.013

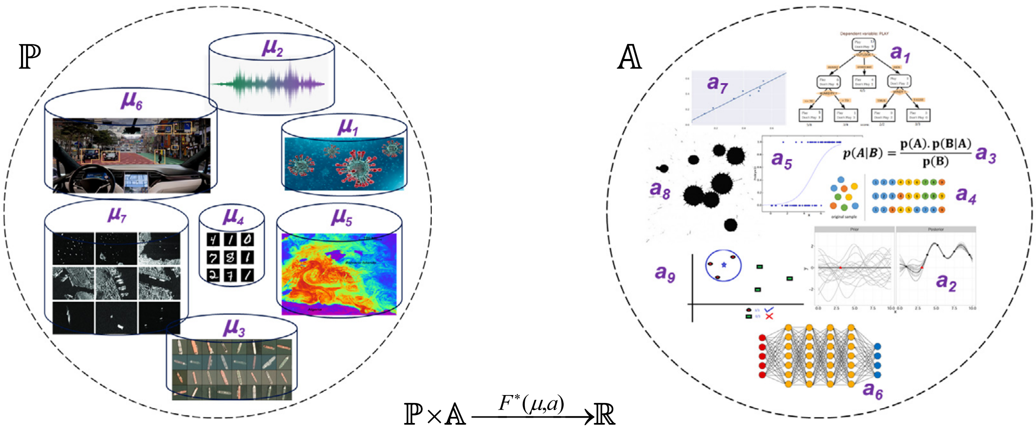

Fig. 1. Different data sets may come from the same probability distribution on a domain. We consider $\mathbb{P}$ to be a compact subset of such distributions. $\mathbb{A}$ is the set of algorithms of interest to us, and F^* is the function that maps a given probability distribution and an algorithm to an m-dimensional vector of quantities of interest.

Intuitively, we wish to obtain a grid on the set of data sets and algorithms, i.e., a finite set of data sets and algorithms so that for every algorithm of interest on every data set of interest, there is some point on the grid that is close to the data set and algorithm, as measured by some parameters. To make this more precise, we clarify what the terms "data sets" and "algorithms" mean for our purposes.

We will assume that each data set is a random sample of an unknown probability distribution in a domain. To be precise, we assume that each distribution is supported on some compact subset of an ambient Euclidean space of dimension q, without loss of generality, on $[-1,1]^q$. Of course, different samples may come from the same distribution, in which case there is no theoretical difference between two such data sets. On the other hand, problems of sample bias are sometimes dealt with by omitting some of the components from each of these samples. Naturally, the resulting data have a different distribution, so the reduced data set is considered in this paper to be a different data set from the original.

In view of the Riesz representation theorem and the Banach-Alaoglu theorem, the set of all probability measures is a compact subset of the dual space $(C([-1,1]^q))^*$. This set is an unmanageably large set representing *every possible* data set that could arise. We model the set of data sets of interest by a smaller compact subset $\mathbb{P}$ of the dual space $(C([-1,1]^q))^*$.

A clear description of the meaning of the term "algorithm" and a precise mathematical definition of the term can be found in [2, Section 1.1]. An algorithm is a function from the input space (the data set) to the output space (real numbers, class labels, etc.) with some additional properties. As in the notion of Rashomon sets as explained in [5,1], one is not interested in the actual algorithms themselves but more in how they perform different tasks on data sets with respect to certain parameters such as stability, accuracy, complexity level of the algorithms, etc. The choice of these parameters will depend upon the applications; for example, accuracy might be more important than speed in numerical applications, while computational time might be more important in time-sensitive applications. It is unlikely that two algorithms will match in terms of all these parameters for all data sets in question. However, if there are two algorithms (or network architectures with different complexity levels) that lead to the same measurements of these quantities, then there is no need to distinguish between them. The stability of an algorithm should mean that when two data sets (meaning two probability distributions) are "close by," then the accuracy and complexity of the algorithm on the two data sets should be close as well. This is captured by the notion of smoothness of the algorithms considered as functions on the data sets.

We assume a set $\mathbb{A}$ of algorithms that act on each data set in $\mathbb{P}$. Each of these algorithms gives rise to a certain number m of parameters. Thus, we are interested in a mapping $F^*:\mathbb{P}\times\mathbb{A}\to\mathbb{R}^m$. Without loss of generality, we may assume $m=1$ in this paper. This is represented in Fig. 1.

We do not expect two algorithms to agree on all the data sets with respect to all of these parameters, that is, we assume that if $a_1, a_2 \in \mathbb{A}$ and

$$F^*(\mu,a_1)=F^*(\mu,a_2) \quad \text{for all } \mu\in\mathbb{P} \Rightarrow a_1=a_2.$$

This means that every $a\in\mathbb{A}$ corresponds to a **unique** mapping F_a on $\mathbb{P}$ defined by

$$F_a(\mu)=F^*(\mu,a), \qquad \mu\in\mathbb{P}. \tag{1.1}$$

An algorithm $a\in\mathbb{A}$ is defined to be *stable* if F_a is a continuous function on $\mathbb{P}$ with a properly defined topology on $\mathbb{P}$.

These considerations prompt us to consider a set $\mathfrak{X}$ of continuous functions from $\mathbb{P}$ to $\mathbb{R}^m$. We will assume implicitly that to every element $F\in\mathfrak{X}$ corresponds a (necessarily unique) algorithm $a\in\mathbb{A}$ such that $F=F_a$ as defined in (1.1). We will then abuse the notation and refer to $F\in\mathfrak{X}$ as an algorithm.

In this paper, we will assume that both $\mathbb{P}$ and $\mathfrak{X}$ are compact metric spaces with appropriate metrics. In fact, in view of the Ascoli theorem, $\mathfrak{X}$ is then an equicontinuous family of functions on $\mathbb{P}$. We then fix a "tolerance" $\epsilon>0$, and find

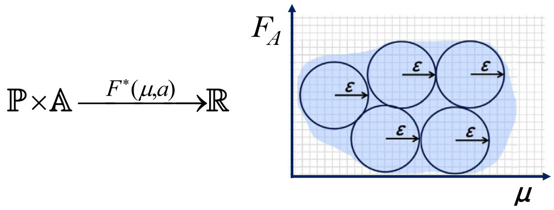

Fig. 2. The ϵ-net for the set $\mathbb{P}\times\mathbb{A}$, where $\mathbb{A}$ is identified with a set of functionals on $\mathbb{P}$. The problem is to estimate the minimum number of balls of radius ϵ to cover the set; the challenge being the fact that $\mathbb{A}$ is a set of functionals acting on an infinite-dimensional space $\mathbb{P}$.

ϵ-nets $\mathbb{P}_\epsilon$ and $\mathfrak{X}_\epsilon$ for $\mathbb{P}$ and $\mathfrak{X}$,[3] respectively. Then $\mathbb{P}_\epsilon\times\mathfrak{X}_\epsilon$ is an ϵ-net for $\mathbb{P}\times\mathfrak{X}$. For any data set $\mu\in\mathbb{P}$ and $F\in\mathfrak{X}$ (equivalently, an algorithm $a\in\mathbb{A}$), there are $\mu_1\in\mathbb{P}_\epsilon$ and $F_1\in\mathfrak{X}_\epsilon$ (equivalently, an algorithm a_1) such that the behavior of a on μ is ϵ-similar to the behavior of a_1 on μ_1. Thus, the problem reduces to finding a minimal ϵ-net for $\mathbb{P}\times\mathfrak{X}$ (or, with our identification of the space $\mathbb{A}$ of algorithms with $\mathfrak{X}$, $\mathbb{P}\times\mathbb{A}$)) as represented in Figs. 1, 2.

The major technical difficulty here is that $\mathfrak{X}$ is a set of functions on an infinite-dimensional space rather than a finite-dimensional Euclidean space as is usual in common machine learning problems. A simplistic idea is to obtain a finite set of parameters for the probability distributions and to treat $\mathfrak{X}$ as a set of functions on these. For example, if $\mathbb{P}$ were a set of normal distributions, then the means and standard deviations would describe this set completely. However, in practice, the distributions are not prescribed in terms of finitely many parameters. Indeed, a central technical challenge in machine learning is that the distributions involved are unknown; in particular, one needs nonparametric methods to deal with these.

It is still possible to restrict ourselves to those distributions that have a smooth density function. In turn, this function can be expanded in an orthogonal series, such as a multivariate tensor product Chebyshev polynomial expansion, and the coefficients of this expansion can be used as parameters for the distribution. If the density functions are smooth enough, then finitely many low-order coefficients will approximate the density well enough, and elements of $\mathfrak{X}$ can be thought of as functions of these low-order coefficients.

Although this simple idea reduces the problem to the case of functions on a Euclidean space, there is still a technical problem. In order to get a good approximation to the density, one needs a large number of coefficients. The curse of dimensionality then poses a big challenge, requiring much more detailed analysis than what is available in the literature.

The organization of this paper is as follows. In Section 2, we review the basic concepts of entropy, analytic, and entire functions. Our main results are stated in Section 3, where we develop an abstract framework, which is then applied to obtain estimates on entropies for certain classes of analytic and entire functions, culminating in the estimates for a class of functionals. In Section 4, we discuss some ideas on how to generate computationally some classes of analytic and entire functions, as well as ϵ-nets for finite-dimensional ellipsoids, which form a theoretical backbone for our estimates. The proofs of the results in Section 3 are given in Section 5. For the convenience of the reader, we include an Appendix, in which we prove certain estimates on the approximation of analytic and entire functions, which motivate our definition of the classes defined in Section 3.

2. Basic concepts

In this section, we explain the basic concepts used in this paper. Section 2.1 describes the multivariate notation. Section 2.2 summarizes the definition of metric entropy and capacity related to the minimal number of balls of a given radius to cover a compact set. The probability measures to be studied have densities that are analytic, while the functionals are entire functions of exponential type defined on an infinite-dimensional sequence space. These ideas are described in Section 2.3. Section 2.4 reviews certain basic notions regarding multivariate Chebyshev polynomials, which are used to encode both analytic and entire functions.

2.1. Multivariate notation

In the sequel, we denote by $d\in\mathbb{N}\cup\{\infty\}$ a generic dimension. Vectors will be denoted by boldface letters, for example, $\mathbf{x}=(x_1,\cdots,x_d)\in\mathbb{R}^d$. The symbol $|\mathbf{x}|_p$ will denote the ℓ^p norm of the vector $\mathbf{x}$. Binary operations among vectors are meant to be in component-wise sense; e.g., $\mathbf{xy}=(x_1y_1,\cdots,x_dy_d)$, $\mathbf{x}^{\mathbf{y}}=\prod_{j=1}^d x_j^{y_j}$, $\mathbf{x}/\mathbf{y}=(x_1/y_1,\cdots,x_d/y_d)$. Similarly, $\mathbf{x}<\mathbf{y}$ means $x_j<y_j$ for $j=1,\cdots,d$, etc. The inner product between two vectors $\mathbf{x},\mathbf{y}$ is denoted by $\mathbf{x}\cdot\mathbf{y}$. For $r>0$, we write $I_r=[-r,r]$, and for a vector $\mathbf{r}$, $I_{\mathbf{r}}=\prod_{j=1}^d[-r_j,r_j]$. Finally, $I=I_1$. For $0<\rho<1$, the ellipse U_ρ is defined by

$$U_\rho=\left\{z\in\mathbb{C}:|z+\sqrt{z^2-1}|<1/\rho\right\},$$

[3] If K is a compact subset of a metric space X and $\epsilon>0$, then a finite set $K_\epsilon\subset X$ is called an ϵ-net for K if K is covered by balls of radius ϵ centered at points in K_ϵ.

where the principal branch of the square root is chosen. With the Joukowski transformation $w = z + \sqrt{z^2 - 1}$, U_ρ is mapped onto the disc $\Gamma_\rho = \{w \in \mathbb{C} : |w| < 1/\rho\}$.

For $0 < \boldsymbol{\rho} = (\rho_1, \ldots, \rho_d) < 1$, the poly-ellipse $U_{\boldsymbol{\rho}}$ (respectively, the poly-disc $\Gamma_{\boldsymbol{\rho}}$) is defined by $U_{\boldsymbol{\rho}} = \prod_{j=1}^{d} U_{\rho_j}$ (respectively, $\Gamma_{\boldsymbol{\rho}} = \prod_{j=1}^{d} \Gamma_{\rho_j}$). When $\boldsymbol{\rho} = (\rho, \cdots, \rho)$, we will abuse the notation and write $U_{\rho,d} = U_{(\rho,\ldots,\rho)}$. If the dimension is clear in the context, we drop the subscript d and write $U_\rho = U_{(\rho,\ldots,\rho)}$. Similar conventions are adopted also for the poly-discs and rectangular cells.

2.2. Entropy and capacity

The material in this section is based on [3, Chapter 15].

Let $(X, \|\cdot\|)$ be a normed linear space, $K \subset X$, and $\epsilon > 0$ be given.

(a) A set $\hat{K} \subset X$ is called an *ϵ-net* for K if, for each $x \in K$, there is at least one $y \in \hat{K}$ such that $\|x - y\| \le \epsilon$.
(b) Points $y_1, \ldots, y_m \in K$ are called *ϵ-separable* if

$$\|y_i - y_j\| \ge \epsilon, \quad i \ne j.$$

Definition 2.1. Let $(X, \|\cdot\|)$ be a normed linear space, and let $K \subset X$ be compact. For any $\epsilon > 0$, let $\mathfrak{N}_\epsilon(K, \|\cdot\|)$ be the minimal value of n such that there exists an ϵ-net for K consisting of n points. The **entropy** of K is defined as

$$H_\epsilon(K, \|\cdot\|) = \log \mathfrak{N}_\epsilon(K, \|\cdot\|). \tag{2.1}$$

Let $\mathfrak{M}_\epsilon(K, \|\cdot\|)$ be the maximum value of m for which there exist m ϵ-separable points for K. The **capacity** of K is defined as

$$C_\epsilon(K, \|\cdot\|) = \log \mathfrak{M}_\epsilon(K, \|\cdot\|). \tag{2.2}$$

The connection between capacity and metric entropy is given in the following proposition.

Proposition 2.1. *Let X be a normed linear space. For each compact set $K \subset X$ and each $\epsilon > 0$,*

$$C_{2\epsilon}(K, \|\cdot\|) \le H_\epsilon(K, \|\cdot\|) \le C_\epsilon(K, \|\cdot\|). \tag{2.3}$$

2.3. Analytic and entire functions

Definition 2.2 *(Analytic functions).* Let $q \in \mathbb{N}$, $\rho > 0$, f be an **analytic function** on $U_\rho := \{\mathbf{z} \in \mathbb{C}^q : \left|z_j + \sqrt{z_j^2 - 1}\right| < 1/\rho,\ j = 1, \ldots, q\}$ if it is complex differentiable at each $\mathbf{z} \in U_\rho$.

Definition 2.3. [Entire functions of exponential type] (a) Let $Q \in \mathbb{N}$, $\tau > 0$. A function $F : \mathbb{C}^Q \to \mathbb{C}$ is called an **entire function of exponential type** τ if

(i) F is an entire function in all of its variables, i.e., F has an absolutely convergent power series expansion

$$F(\mathbf{z}) = \sum_{\mathbf{k} \in \mathbb{N}^Q} a_{\mathbf{k}} \mathbf{z}^{\mathbf{k}}, \qquad \mathbf{z} \in \mathbb{C}^Q$$

with constant coefficients $a_{\mathbf{k}} \in \mathbb{C}$.

(ii) For any $\epsilon > 0$ there exists a positive number A_ϵ such that for all $\mathbf{z} \in \mathbb{C}^Q$, the inequality

$$|F(\mathbf{z})| \le A_\epsilon \exp\left((\tau + \epsilon) \sum_{j=1}^{Q} |z_j|\right)$$

is satisfied.

(b) If $\mathbf{v} = (v_1, \ldots, v_Q) \in \mathbb{R}_+^Q$, then F is said to be an **entire function of exponential type v** if the function $\mathbf{z} \mapsto F(z_1/v_1, \cdots, z_Q/v_Q)$ is an entire function of exponential type 1.
(c) Let $\mathbf{v} \in \ell^1(\mathbb{C})$. A function $F : c_0(\mathbb{C}) \to \mathbb{C}$ is called an **entire function of exponential type v** if, for every $Q \in \mathbb{N}$, the function $(z_1, \cdots, z_Q) \to F(z_1, \cdots, z_Q, 0, 0, \cdots)$ is an entire function of exponential type $(v_1, \cdots, v_Q)$.

An important example of entire functions of finite exponential type on $c_0(\mathbb{C})$ is the mapping

$$\mathbf{z} \in c_0(\mathbb{C}) \mapsto \int \exp(-i\mathbf{z}\cdot\mathbf{x})d\mu(\mathbf{x}),$$

where μ is a probability measure supported on the infinite cube $[-1,1]^\infty$.

2.4. Chebyshev polynomials

Let $d \in \mathbb{N}$.

$$v_d(\mathbf{x}) = \pi^{-d}\prod_{k=1}^{d}(1-x_k^2)^{-1/2}, \qquad \mathbf{x} = (x_1,\cdots,x_d) \in I^d. \tag{2.4}$$

The space $L^p(I^d)$ will refer to the space of all f for which

$$\|f\|_{d,p} = \begin{cases} \left\{\displaystyle\int_{I^d} |f(\mathbf{x})|^p v_d(\mathbf{x})d\mathbf{x}\right)^{1/p}, & \text{if } 0<p<\infty,\\ \displaystyle\operatorname*{ess\,sup}_{\mathbf{x}\in I^d}|f(\mathbf{x})|, & \text{if } p=\infty, \end{cases} \tag{2.5}$$

is finite. As usual, we will identify two functions if they are equal almost everywhere.

We denote the space of all polynomials in d variables of coordinatewise degree $<n$ by Π_n^d.

Next, we define the Chebyshev polynomials. We define Chebyshev polynomials in the univariate case by first setting $x=\cos\theta$ for $x\in[-1,1]$ and define

$$p_k(x) = \begin{cases} 1, & \text{if } k=0,\\ \sqrt{2}\cos(k\theta), & \text{if } k=1,2,\cdots. \end{cases} \tag{2.6}$$

We note that the expression p_k is a polynomial of degree k in x, and the normalization is set so that

$$\int_{-1}^{1} p_k(x)p_j(x)v_1(x)dx = \delta_{k,j}. \tag{2.7}$$

The multivariate Chebyshev polynomials are defined by

$$p_{\mathbf{k}}(\mathbf{x}) = \prod_{j=1}^{d} p_{k_j}(x_j), \qquad \mathbf{k}=(k_1,\cdots,k_d)\in\mathbb{Z}_+^d,\ \mathbf{x}=(x_1,\cdots,x_d)\in I^d, \tag{2.8}$$

and satisfy

$$\int_{I^d} p_{\mathbf{k}}(\mathbf{x})p_{\mathbf{j}}(\mathbf{x})v_d(\mathbf{x})d\mathbf{x} = \delta_{\mathbf{k},\mathbf{j}}. \tag{2.9}$$

We note that even though we have defined the Chebyshev polynomials by their values on I^d, they are actually defined on $\mathbb{C}^d$ because they are polynomials.

Any function $f \in L^2(I^d)$ admits a formal expansion

$$f = \sum_{\mathbf{k}\in\mathbb{N}^d} \hat{f}(\mathbf{k})p_{\mathbf{k}}, \tag{2.10}$$

where the **Chebyshev coefficients** are defined by

$$\hat{f}(\mathbf{k}) = \int_{I^d} f(\mathbf{y})p_{\mathbf{k}}(\mathbf{y})v_d(\mathbf{y})d\mathbf{y}, \qquad \mathbf{k}\in\mathbb{Z}_+^d. \tag{2.11}$$

For $f \in L^1(I^d)$, we define the partial sums of (2.10) by

$$s_n(f)(\mathbf{x}) = \sum_{|\mathbf{k}|_1<n} \hat{f}(\mathbf{k})p_{\mathbf{k}}(\mathbf{x}), \qquad S_n(f)(\mathbf{x}) = \sum_{|\mathbf{k}|_1=n} \hat{f}(\mathbf{k})p_{\mathbf{k}}(\mathbf{x}), \qquad n\in\mathbb{N}. \tag{2.12}$$

There is an important formula that relates Chebyshev expansions to Laurent expansions of meromorphic functions. We note that for $0 < \rho < 1$, the two branches of the Joukowski transform $w = z + \sqrt{z^2 - 1}$ map U_ρ to the annulus $\rho < |w| < 1/\rho$. Hence, for a function f analytic on U_ρ for some $\rho > 0$, the function $g(w) = f((w + w^{-1})/2)$ is analytic on the annulus. The Laurent expansion of g is given by

$$\hat{f}(0) + (1/2)\sum_{k=1}^{\infty} \hat{f}(k)(w^k + w^{-k}). \tag{2.13}$$

Thus, the coefficients, the partial sums, and the remainder $f - s_n(f)$ can be expressed as a contour integral on the appropriate circles in the w plane. For multivariate functions, of course, one uses the tensor products of circles.

If $\boldsymbol{r} > 0$, we define Chebyshev polynomials on $I_{\boldsymbol{r}}$ by

$$p_{\mathbf{k},\boldsymbol{r}}(\mathbf{x}) = \prod_{j=1}^{d} p_{k_j}(\mathbf{x}/\boldsymbol{r}), \tag{2.14}$$

and the corresponding weights by

$$v_{\boldsymbol{r}}(\mathbf{x}) = \pi^{-d}\left(\prod_{j=1}^{d}(r_j^2 - x_j^2)\right)^{-1/2}. \tag{2.15}$$

Of course, one has the orthogonality relation

$$\int_{I_{\boldsymbol{r}}} p_{\mathbf{k},\boldsymbol{r}}(\mathbf{x}) p_{\mathrm{m},\boldsymbol{r}}(\mathbf{x}) v_{\boldsymbol{r}}(\mathbf{x}) d\mathbf{x} = \delta_{\mathbf{k},\mathrm{m}}. \tag{2.16}$$

The Chebyshev coefficients and partial sums are defined in an obvious way and will be indicated by an extra subscript $\boldsymbol{r}$; for example, $s_{n,\boldsymbol{r}}$.

3. Main results

In this section, we define compact spaces of analytic and entire functions and state our theorems about their entropies. In Section 3.1, we encapsulate the procedure in some abstraction. The spaces for analytic functions and their entropy estimates are given in Section 3.2. Analogous results for entire functions are given in Section 3.3. We conclude with estimates on the entropy of functionals in Section 3.4.

3.1. Direct sums and products

Let X be a Banach space. We assume that there exists a sequence of finite-dimensional subspaces X_j, $j = 0, 1, \ldots$, $b_j = \mathrm{dim}(X_j)$, $Y_k = \bigoplus_{j=0}^{k-1} X_j$, $d_k = \mathrm{dim}(Y_k) = \sum_{j=0}^{k-1} b_j$, such that $\bigcup_{k=0}^{\infty} Y_k$ is dense in X. In particular, we assume that for any $f \in X$, there is a unique sequence $\{f_j \in X_j\}_{j=0}^{\infty}$ such that we have a formal expansion of the form $f \sim \sum_j f_j$. (An example is that the space $X = C(I^q)$, X_j be the space Π_j^q of q-variate polynomials of degree $< j$, and $f_j = S_{j-1}(f)$, as in (2.12).) We write $\mathrm{Proj}_j(f) = f_j$ and assume that Proj_j is a continuous operator for each j. Generalizing the notation established in Section 2.4, we define

$$s_n(f) = \sum_{j=0}^{n-1} \mathrm{Proj}_j(f), \quad S_k(f) = \mathrm{Proj}_k(f).$$

Let $\mathfrak{K}$ be a compact subset of X. Then

$$\lim_{n\to\infty} \sup_{f\in\mathfrak{K}} \mathrm{dist}(f, Y_n) = 0.$$

In this paper, we are interested in $\mathfrak{K}$ such that

$$\lim_{n\to\infty} \sup_{f\in\mathfrak{K}} \|f - s_n(f)\| = 0. \tag{3.1}$$

More precisely, with a summable sequence $\{\Delta_j\}_{j=0}^{\infty}$ of positive numbers, we define

$$\mathfrak{K} = \left\{ f \in X : \|S_j(f)\| \le \Delta_j,\ j \in \mathbb{N} \right\}, \qquad \mathfrak{K}_j = \mathrm{Proj}_j(\mathfrak{K}), \qquad \widetilde{\mathfrak{K}}_n = \bigoplus_{j=0}^{n-1} \mathfrak{K}_j. \tag{3.2}$$

Let $\epsilon > 0$. To estimate the entropy of $\mathfrak{K}$, we first observe that, in view of (3.1), there exists some $n \in \mathbb{N}$ such that

$$\sup_{f \in \mathfrak{K}} \|f - s_n(f)\| \le \epsilon/2.$$

Thus, any $\epsilon/2$-net of the set $\widetilde{\mathfrak{K}}_n$ is an ϵ-net of $\mathfrak{K}$, and any ϵ-net of the set $\mathfrak{K}$ is an ϵ-net of $\widetilde{\mathfrak{K}}_n$. Thus,

$$H_\epsilon\left(\widetilde{\mathfrak{K}}_n, X\right) \le H_\epsilon(\mathfrak{K}, X) \le H_{\epsilon/2}\left(\widetilde{\mathfrak{K}}_n, X\right). \tag{3.3}$$

Therefore, in order to estimate the entropy of $\mathfrak{K}$, we only need to estimate the entropy of $\widetilde{\mathfrak{K}}_n$.

For this purpose, it is convenient to identify $\widetilde{\mathfrak{K}}_n$ with a tensor product of balls.

We consider the space $X_{\Pi,n} = \prod_{j=0}^{n-1} X_j$, and the mapping $\mathcal{T}_n(f) = (\mathrm{Proj}_0(f), \cdots, \mathrm{Proj}_{n-1}(f))$ from $\bigoplus_{j=0}^{n-1} X_j$ to $X_{\Pi,n}$. Obviously, $\mathcal{T}_n$ is a one-to-one mapping. If $1 \le p \le \infty$, we may define a norm on $X_{\Pi,n}$ by

$$\|\mathcal{T}_n(f)\|_{\Pi,p,n} = \left|(\|\mathrm{Proj}_0(f)\|, \dots, \|\mathrm{Proj}_{n-1}(f)\|)\right|_p. \tag{3.4}$$

Since all the spaces involved are finite dimensional, there exist positive constants $A_{n,p}, B_{n,p}$ such that

$$A_{n,p}\|f\| \le \|\mathcal{T}_n(f)\|_{\Pi,p,n} \le B_{n,p}\|f\|, \quad f \in Y_n. \tag{3.5}$$

Next, we note that $\mathfrak{K}_j$ is a ball in finite dimensional space X_j:

$$\mathfrak{K}_j = \left\{ f \in X_j : \|f\| \le \Delta_j \right\}, \quad j \in \mathbb{N}. \tag{3.6}$$

So, we can view $\widetilde{\mathfrak{K}}_n$ through the mapping $\mathcal{T}_n$ as a product of the balls $\mathfrak{K}_j$. The entropy of this product is given in [3, Proposition 1.3]. To summarize, the entropy of $\mathfrak{K}$ can be estimated as in the following theorem.

Theorem 3.1. *Let $\epsilon > 0$, $1 \le p, r \le \infty$, and we recall the notation established in (3.5), (3.6). We have*

$$\sum_{j=0}^{N-1} b_j \log\left(\frac{\Delta_j}{2B_{N,r}\epsilon}\right) \le H_\epsilon(\mathfrak{K}) \le \sum_{j=0}^{M-1} b_j \log\left(\max\left(\frac{6M^{1/p}\Delta_j}{A_{M,p}\epsilon}, 1\right)\right), \tag{3.7}$$

which holds for all $N \ge 1$ and

$$M \ge \mathcal{N}(\epsilon/2) := \min\left\{ m \in \mathbb{N} : \sum_{n=m}^{\infty} \Delta_n \le \epsilon/2 \right\}. \tag{3.8}$$

3.2. Spaces of analytic functions

Let $q \in \mathbb{N}$, $\rho \in (0,1)$. In view of Theorem A.1, we define the class of analytic functions by

$$\mathcal{A}_\rho = \left\{ f : I^q \to \mathbb{R} : \|S_n(f)\|_{L^2(I^q)} \le \rho^n, \quad n \in \mathbb{N} \right\}. \tag{3.9}$$

The goal of this section is to prove Theorem 3.2 to estimate the entropy of $\mathcal{A}_\rho$.

We will use Theorem 3.1 with $A_{n,2} = B_{n,2} = 1$, $\Delta_j = \rho^j$, $b_j = \binom{j+q-1}{q-1}$, $j = 0, 1, \dots$, to obtain the following theorem.

Theorem 3.2.

(a) *For*

$$\epsilon < \min\left\{ \frac{2}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2 \log\frac{1}{\rho}}{9(1-\rho^2)}\right), \frac{2\rho}{\sqrt{1-\rho^2}} \left(\frac{2\log\frac{1}{\rho}}{9(\rho^{-2}-1)(q+1)}\right)^{q+1} \right\}, \tag{3.10}$$

we have

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \le \frac{4e^{q+1}}{\sqrt{2\pi}} \left(1 + \frac{\log\left(\frac{2\rho}{\sqrt{1-\rho^2}}\frac{1}{\epsilon}\right)}{(q+1)\log\frac{1}{\rho}}\right)^{q+1} \log\frac{1}{\rho} \tag{3.11}$$

(b) *For $\epsilon < 1/2$, we have*

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \geq \frac{2^{q+1}}{8\sqrt{2\pi(q+1)}}\left(1+\frac{\log\left(\frac{\rho^2}{4}\frac{1}{\epsilon}\right)}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}\log\frac{1}{\rho}. \tag{3.12}$$

(c) *Furthermore, for ϵ sufficiently small, the entropy of $\mathcal{A}_\rho$ satisfies*

$$1-\frac{2(q+1)\log\frac{2}{\rho}}{\log\frac{1}{\epsilon}} \leq \frac{H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)})}{\frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}} \leq 1+\frac{2(q+1)\log\frac{1}{\rho}}{\log\frac{1}{\epsilon}}\left(\log\log\frac{1}{\epsilon}+\frac{\log\frac{2}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}}+q+4\right). \tag{3.13}$$

3.3. Spaces of entire functions

In this section, we are interested in the class of entire functions of finite exponential type, defined in (3.16) below. We will use Theorem 3.1 again to estimate the entropy of this class. The main difficulty in this section is keeping track of the dependence of the dimension Q. This is important when we consider functional classes in Section 3.4.

Let $Q \in \mathbb{N}$,

$$1 \leq \tau\frac{Q}{2e^{3/2}\pi}, \qquad C=\left(\frac{2\pi}{Q}\right)^{Q/2}, \tag{3.14}$$

and let

$$\Lambda(N) = CN^{Q/2}\frac{\tau^N}{N!}, \quad N \in \mathbb{N}. \tag{3.15}$$

For $\mathbf{r} \in \mathbb{R}^Q_+$, let $I_{\mathbf{r}} = \prod_{j=1}^{Q}[-r_j, r_j]$ be a subset of $\mathbb{R}^Q$. In view of Theorem A.2 and A.3, we can define the class of entire functions by

$$\mathcal{B}_Q = \mathcal{B}_Q(\mathbf{r}, \tau) = \left\{F : I_{\mathbf{r}} \to \mathbb{R} : \|S_N(F)\|_{L_\infty(I_{\mathbf{r}})} \leq \Lambda(N),\ N \in \mathbb{N}\right\}. \tag{3.16}$$

Theorem 3.3. *Let (3.14) hold.*

(a) *Under the condition that*

$$\epsilon \leq \left(\frac{2\pi e\tau}{Q}\right)^{Q/2}\frac{4}{(e\tau)^{1/2}\exp(e^2\tau)}, \tag{3.17}$$

the entropy of $\mathcal{B}_Q$ defined in (3.16) satisfies

$$\begin{aligned}&H_\epsilon(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})})\\ &\leq \frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}+\frac{3Q}{4}\right)^{Q+1}\left(7\log\log\frac{1}{\epsilon}+\log\left((Q+1)^2(e\tau)^6\right)\right).\end{aligned} \tag{3.18}$$

(b) *Let*

$$\xi_\tau = \frac{16\max\{3e^2\tau, 128\}\log\left(\max\{3e^2\tau, 128\}\right)}{e}+2,$$

under the condition that

$$\epsilon \leq \left(\frac{2\pi}{Q}\right)^{Q/2}\frac{1}{4\sqrt{2\pi e\tau}}\xi_\tau^{-2\xi_\tau}, \tag{3.19}$$

$$\begin{aligned}H_\epsilon(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}) \geq &\frac{1}{16\sqrt{\pi Q}}\left(\frac{1}{Q}\right)^Q\left(\frac{\log\frac{1}{4\sqrt{2\pi e\tau\epsilon}}+\frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)}{\log\left(\log\frac{1}{4\sqrt{2\pi e\tau\epsilon}}+\frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)\right)-\log(e\tau)}-\frac{5}{2}+2Q\right)^Q\\ &\times\left(\frac{\log\frac{1}{4\sqrt{2\pi e\tau\epsilon}}+\frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)}{\log\left(\log\frac{1}{4\sqrt{2\pi e\tau\epsilon}}+\frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)\right)-\log(e\tau)}-\log(e\tau)-\frac{3}{2}\right).\end{aligned} \tag{3.20}$$

(c) *The following asymptotic result holds:*

$$\frac{1}{2Q!}\frac{\left(\log\frac{1}{\epsilon}\right)^{Q+1}}{\left(2\log\log\frac{1}{\epsilon}\right)^{Q}}(1+o(1)) \le H_\epsilon(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}) \le \frac{1}{Q!}\frac{\left(2\log\frac{1}{\epsilon}\right)^{Q+1}}{\left(\log\log\frac{1}{\epsilon}\right)^{Q}}(1+o(1)) \tag{3.21}$$

as $\epsilon \to 0$, *where the* $o(1)$ *term is* $\sim Q\log\log(1/\epsilon)/\log(1/\epsilon)$.

3.4. Space of functionals

In this section, we are interested in estimating the entropy of a class of functionals $\mathcal{F}$ on $\mathcal{A}_\rho$ with respect to the sup-norm. Any functional in $\mathcal{F}$ can be viewed as a functional on the sequence of Chebyshev coefficients of the input function. We will define $\mathcal{F}$ to be a set of functionals that are entire functions of certain exponential type as in Definition 2.3(c).

Under the assumption that the functionals are Lipschitz continuous, i.e.,

$$\sup_{\tilde{F}\in\mathcal{F}}\ \sup_{f_1\neq f_2\in\mathcal{A}_\rho}\frac{|\tilde{F}(f_1)-\tilde{F}(f_2)|}{\|f_1-f_2\|_{L^2(I^q)}} \le 1,$$

we conclude for any $\epsilon > 0$, there is some integer n such that for any $\tilde{F}\in\mathcal{F}$,

$$\left|\tilde{F}(f)-\tilde{F}(s_{n+1}(f))\right| \le \|f-s_{n+1}(f)\|_{L^2(I^q)} \le \epsilon/2, \qquad \mathcal{A}_\rho.$$

Consequently, the ϵ-entropy of $\mathcal{F}$ is bounded by the $\epsilon/2$-entropy of

$$\left\{\tilde{F}\circ s_{n+1} :\ \tilde{F}\in\mathcal{F}\right\}.$$

In turn, for any $f\in\mathcal{A}_\rho$, $\tilde{F}\circ s_{n+1}(f)$ can be viewed as a function of the Chebyshev coefficients of f up to order $n+1$.

We now define the set of functionals formally. Let $q\in\mathbb{N}$, $\rho<1$, $\mathcal{A}_\rho$ be as in Theorem 3.2. For any $n\in\mathbb{N}$, write $Q=Q(n)=\binom{n+q}{q}$. The distance $\|f-s_{n+1}(f)\|_{L^2(I^q)}$ is bounded as

$$\|f-s_{n+1}(f)\|_{L^2(I^q)} \le \left(\sum_{j=n+1}^{\infty}\rho^{2j}\right)^{1/2} \le \frac{\rho^{n+1}}{\sqrt{1-\rho^2}}.$$

By definition, the Chebyshev coefficients of $f\in\mathcal{A}_\rho$ satisfy $|\hat{f}(\mathbf{k})|\le\rho^{|\mathbf{k}|_1}$. Let $\tilde{\mathbf{r}}=(r_j)_{j=1}^\infty$ be defined by

$$r_j=\rho^\ell, \qquad \text{for } \binom{q+\ell-1}{q} < j \le \binom{q+\ell}{q}$$

and let $\mathbf{r}=(r_j)_{j=1}^Q$. Then, for $f\in\mathcal{A}_\rho$, $\left(\hat{f}(\mathbf{k})\right)_{|\mathbf{k}|_1\le n}\in I_{\mathbf{r}}\subset\mathbb{R}^Q$. Consequently, the functionals on the polynomial space Π_{n+1}^q are identified as functions on $I_{\mathbf{r}}$ as follows:

$$F\left((a_{\mathbf{k}})_{|\mathbf{k}|_1\le n}\right)\mapsto\tilde{F}\left(\sum_{\mathbf{k}_1\le n}a_{\mathbf{k}}p_{\mathbf{k},\mathbf{r}}\right).$$

The functionals on $\mathcal{A}_\rho$ with which we are concerned are functionals that induce entire functions of some type $\mathbf{v}$ by this process.

Let $\tilde{\mathbf{v}}=(v_j)_{j=1}^\infty$ be a nonnegative sequence. For any $n\in\mathbb{N}$, let $Q=\binom{n+q}{q}$, $\mathbf{r}=(r_j)_{j=1}^Q$; we denote the class of functionals $\mathcal{F}_{n,\tilde{\mathbf{v}}}$ on $\mathcal{A}_\rho$ by

$$\mathcal{F}_{n,\tilde{\mathbf{v}}} := \left\{\tilde{F}:\ \mathcal{A}_\rho\to\mathbb{R}:\ \exists F\in\tilde{\mathcal{B}}_{n,\tilde{\mathbf{v}}} \text{ such that } \tilde{F}(f)=F\left(\left(\hat{f}(\mathbf{k})\right)_{|\mathbf{k}|_1\le n}\right),\ f\in\mathcal{A}_\rho\right\}, \tag{3.22}$$

where each $\tilde{\mathcal{B}}_{n,\tilde{\mathbf{v}}}$ is denoted as

$$\tilde{\mathcal{B}}_{n,\tilde{\mathbf{v}}} = \left\{F: I_{\mathbf{r}}\to\mathbb{R}:\ \|S_N(F)\|_{L_\infty(I_{\mathbf{r}})} \le \left(\frac{2\pi}{Q}\right)^{Q/2}N^{Q/2}\frac{1}{N!}\left(\sum_{j=1}^{Q}v_jr_j\right)^N\right\}.$$

We denote the class of functionals on $\mathcal{A}_\rho$ in this section $\mathcal{F}_{\tilde{\mathbf{v}}}$ as

$$\mathcal{F}_{\tilde{\mathbf{v}}} = \left\{\tilde{F}:\ \mathcal{A}_\rho\to\mathbb{R}:\ \sup_{f_1\neq f_2}\frac{|\tilde{F}(f_1)-\tilde{F}(f_2)|}{\|f_1-f_2\|_{L^2(I^q)}}\le 1,\ \tilde{F}\circ s_{n+1}\in\mathcal{F}_{n,\tilde{\mathbf{v}}},\quad \forall n\in\mathbb{N}\right\} \tag{3.23}$$

and the metric on $\mathcal{F}_{\tilde{\mathbf{v}}}$ is

$$\|F\|_{\mathcal{F}_{\tilde{\mathbf{v}}}} = \sup_{f\in\mathcal{A}_\rho} |F(f)|, \quad F \in \operatorname{span}(\mathcal{F}_{\tilde{\mathbf{v}}}).$$

We estimate the entropy of the class $\mathcal{F}_{\tilde{\mathbf{v}}} \times \mathcal{A}_\rho$, with respect to the metric $\|\cdot\|$ defined by

$$\|\cdot\| = \|\cdot\|_{L^2(I^q)} + \|\cdot\|_{\mathcal{F}_{\tilde{\mathbf{v}}}}.$$

Our main theorem in this subsection gives a bound of the entropy $H_\epsilon\left(\mathcal{F}_{\tilde{\mathbf{v}}} \times \mathcal{A}_\rho, \|\cdot\|\right)$.

Theorem 3.4. *Let $\tilde{\mathbf{v}} = (v_j)_{j=1}^\infty$ be denoted by*

$$v_j = \frac{1}{2e^{3/2}\pi\rho^\ell}, \qquad \text{for } \binom{q+\ell-1}{q} < j \le \binom{q+\ell}{q}, \tag{3.24}$$

then for

$$\epsilon < \min\left\{ \frac{2}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2\log\frac{1}{\rho}}{9(1-\rho^2)}\right), \frac{2\rho}{\sqrt{1-\rho^2}}\left(\frac{2\log\frac{1}{\rho}}{9(\rho^{-2}-1)(q+1)}\right)^{q+1}, \frac{4\rho^q}{\sqrt{1-\rho^2}} \right\}, \tag{3.25}$$

the entropy of $\mathcal{F}_{\tilde{\mathbf{v}}} \times \mathcal{A}_\rho$ is bounded by

$$H_\epsilon(\mathcal{F}_{\tilde{\mathbf{v}}} \times \mathcal{A}_\rho, \|\cdot\|) \le \frac{60q}{3\sqrt{2\pi}} \exp\left\{\gamma^q \log\left(\frac{e^{3/2}}{\pi} + 2e\right)\right\} \gamma^q \log(q\gamma\log\gamma) + \gamma^{q+1}\log\frac{1}{\rho} \tag{3.26}$$

with

$$\gamma = \frac{2e\log\frac{1}{\epsilon}}{q\log\frac{1}{\rho}}. \tag{3.27}$$

Remark 3.1. At a first glance, the lower bound for the entropy of the set $\mathcal{F}_{n,\tilde{\mathbf{v}}}$ can be derived immediately from Theorem 3.3. However, in the definition of $\mathcal{F}_{\tilde{\mathbf{v}}}$, there is a Lipschitz condition under which we can only consider subsets of $\mathcal{F}_{n,\tilde{\mathbf{v}}}$ in the proof. These subsets do not fit our abstract framework. Therefore, we are not able to obtain a lower bound at this time. □

4. Computational issues

4.1. Generating analytic and bandlimited functions

A simple way to generate functions that are analytic on the interior of the poly-ellipse $U_\rho \subset \mathbb{C}^d$:

$$\mathbf{x} = \mathbf{x}(\boldsymbol{\theta}) = \frac{\rho+\rho^{-1}}{2}\cos(\boldsymbol{\theta}), \quad \mathbf{y} = \mathbf{y}(\boldsymbol{\theta}) = \frac{\rho-\rho^{-1}}{2}\sin(\boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in (-\pi,\pi]^d, \tag{4.1}$$

is the following. We take a random sample $\{\boldsymbol{\theta}_j\}_{j=1}^M$ on $(-\pi,\pi]^d$ and generate points $\mathbf{w}_j = \mathbf{x}(\boldsymbol{\theta}_j) + i\mathbf{y}(\boldsymbol{\theta}_j)$ on U_ρ. We also take a random sample $\{a_j\}_{j=1}^M$ from some compact subset of $\mathbb{C}$. Then the function

$$f(\mathbf{z}) = \sum_{j=1}^M \frac{a_j}{\mathbf{w}_j - \mathbf{z}} + \sum_{j=1}^M \frac{\overline{a_j}}{\overline{\mathbf{w}_j} - \mathbf{z}}$$

is clearly analytic in the interior of U_ρ and real-valued on $[-1,1]^d$. A probability density on $[-1,1]^d$ can be obtained by normalizing $f(\mathbf{x})^2$ to have integral equal to 1. Different choices of the random samples yield different distributions.

To generate band-limited functions on $c_0(\mathbb{C})$, we use a similar idea. We consider random sequences $\mathbf{w}_j \in \ell^1$, and samples $\{a_j\}$ on a complex ellipsoid, $j = 1, \cdots, M$. For any such sequence and random sample, we have a band-limited function of the form

$$f(\mathbf{z}) = \sum_{j=1}^M a_j \exp(i\mathbf{z}\cdot\mathbf{w}_j) + \sum_{j=1}^M \overline{a_j}\exp(-i\mathbf{z}\cdot\overline{\mathbf{w}_j}),$$

which are real-valued for real sequences $\mathbf{z}$.

4.2. Generating ϵ-nets on ellipsoids

We note first that for any norm $\|\cdot\|$ on $\mathbb{R}^d$, the ellipsoid $\mathbb{B}(\mathbf{x}_0, \boldsymbol{r})$ centered at $\mathbf{x}_0$ is parametrized by

$$x_j = x_{0,j} + r_j y_j,$$

where $\mathbf{y} = (y_j)$ belongs to the unit ball $\mathbb{B}(\mathbf{0}, 1)$. Therefore, it is enough to generate a net for this ball; the net on the ellipsoid can be generated by appropriate scaling. Accordingly, we describe the generation of an ϵ-net for $\mathbb{B}(\mathbf{0}, 1)$.

In [4, Proof of Lemma 7.1], we have proved that if $\delta \in (0, 1)$,

$$M \geq (4/\epsilon)^d \log\left(\frac{(12/\epsilon)^d}{\delta}\right),$$

and $\mathcal{C} = \{\mathbf{z}_1, \cdots, \mathbf{z}_M\}$ is a random sample from the uniform distribution on $\mathbb{B}(\mathbf{0}, 1)$, then with probability exceeding $1 - \delta$, $\mathcal{C}$ is an $\epsilon/2$-net for $\mathbb{B}(\mathbf{0}, 1)$. To find a minimal ϵ-net, we use a greedy algorithm: start with $\mathcal{C} = \{z_1\}$, and for $j = 2, \cdots, M$, add the point $\mathbf{z}_j$ to $\mathcal{C}$ if $\mathsf{dist}(\mathcal{C}, \mathbf{z}_j) \geq \epsilon/2$. Then clearly, $\mathcal{C}$ is an $\epsilon/2$-separated subset and ϵ-net of $\mathbb{B}(\mathbf{0}, 1)$.

5. Proofs

This section is organized as follows. In Section 5.1, we introduce some basic lemmas on binomial coefficients, which are used multiple times in the rest of the proof. In Section 5.2, we prove Theorem 3.1. This theorem is an abstract theorem, which can be applied to prove the entropy of analytic and entire function classes. Section 5.3 is the proof of Theorem 3.2. Section 5.4 and Section 5.5 are the proof of Theorem 3.3. Section 5.6 is the proof of Theorem 3.4, which shows the entropy of functional classes defined in Section 3.4.

5.1. Combinatorial identities and inequalities

Lemma 5.1. *Let $n, d \in \mathbb{N}$, $n \geq 1$. Then we have*

$$\sum_{j=0}^{n} \binom{j+d-1}{d-1} = \binom{n+d}{d}, \tag{5.1}$$

$$\sum_{j=0}^{n} j\binom{j+d-1}{d-1} = d\binom{n+d}{d+1}, \qquad \sum_{j=1}^{n}(n-j)\binom{j+d-1}{d-1} = \binom{n+d}{d+1}. \tag{5.2}$$

Proof of Lemma 5.1. The identity (5.1) follows by noticing that $\binom{d-1}{d-1} = \binom{d}{d} = 1$ and

$$\begin{aligned}\binom{j-1+d}{d} + \binom{j+d-1}{d-1} &= \frac{(n+d-1)\ldots n}{d!} + \frac{(n+d-1)\ldots(n+1)}{(d-1)!} \\ &= \frac{(n+d-1)\ldots(n+1)}{(d-1)!}\left(1+\frac{n}{d}\right) = \frac{(n+d)\ldots(n+1)}{d!} \\ &= \binom{j+d}{d}.\end{aligned}$$

The first identity in (5.2) is given by

$$\begin{aligned}&\sum_{j=0}^{n} j\binom{j+d-1}{d-1} = \sum_{j=0}^{n} j\frac{(j+d-1)\ldots(j+1)}{(d-1)!} = \sum_{j=0}^{n}\frac{(j+d-1)\ldots(j+1)j}{d!}d \\ &= \sum_{j=0}^{n} d\binom{j+d-1}{d} = d\binom{n+d}{d+1}.\end{aligned} \tag{5.3}$$

The second identity in (5.2) is a simple calculation using the first identity and (5.1). □

Lemma 5.2. *Let $n, d \in \mathbb{N}$, then*

$$\frac{1}{8\sqrt{2\pi(d+1)}}\left(\frac{2(n+d)}{d+1}\right)^{d+1} \leq \binom{n+d}{d+1} \leq \frac{2}{\sqrt{2\pi}}\left(\frac{e(n+d)}{d+1}\right)^{d+1}. \tag{5.4}$$

In particular, if n has the form

$$n_1 = \lfloor a(h+b) \rfloor, \quad n_2 = \lfloor a(h-b) \rfloor \tag{5.5}$$

for some constants $a > 0$, $b \geq 0$ *and* $h \geq \left(b + \frac{d}{a}\right)(d+1)$, *then*

$$\begin{aligned}\binom{n_1+d}{d+1} &\leq \frac{(ah)^{d+1}}{(d+1)!}\left(1 + \frac{2(d+1)(ab+d)}{ah}\right)\\ \binom{n_2+d}{d+1} &\geq \frac{(ah)^{d+1}}{(d+1)!}\left(1 - \frac{2(d+1)(ab+1)}{ah}\right).\end{aligned} \tag{5.6}$$

Proof of Lemma 5.1. By Stirling's approximation formula,

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \leq k! \leq 2\sqrt{2\pi k}\left(\frac{k}{e}\right)^k.$$

This gives

$$\begin{aligned}&\frac{1}{4}\left(1+\frac{d+1}{n-1}\right)^{n-1}\left(\frac{n-1}{d+1}+1\right)^{d+1}\sqrt{\frac{n+d}{2\pi(d+1)(n-1)}} \leq \binom{n+d}{d+1}\\ \leq &2\left(1+\frac{d+1}{n-1}\right)^{n-1}\left(\frac{n-1}{d+1}+1\right)^{d+1}\sqrt{\frac{n+d}{2\pi(d+1)(n-1)}}\end{aligned} \tag{5.7}$$

and hence

$$\frac{1}{8\sqrt{2\pi(d+1)}}\left(\frac{2(n+d)}{d+1}\right)^{d+1} \leq \binom{n+d}{d+1} \leq \frac{2}{\sqrt{2\pi}}\left(\frac{e(n+d)}{d+1}\right)^{d+1}.$$

Suppose now

$$n_1 = \lfloor a(h+b) \rfloor,$$

then

$$\binom{n_1+d}{d+1} \leq \frac{(ah+ab+d)\dots(ah+ab)}{(d+1)!} \leq \frac{(ah)^{d+1}}{(d+1)!}\left(1+\frac{ab+d}{ah}\right)^{d+1}.$$

Since $h \geq \left(b + \frac{d}{a}\right)(d+1)$, we have

$$\frac{ab+d}{ah} \leq \frac{1}{d+1},$$

then

$$\binom{n_1+d}{d+1} \leq \frac{(ah)^{d+1}}{(d+1)!}\left(1+\frac{2(d+1)(ab+d)}{ah}\right).$$

This proves the upper bound in (5.6).

On the other hand, $h \geq \left(b + \frac{1}{a}\right)(d+1)$, hence

$$\binom{n_2+d}{d+1} \geq \frac{(ah-ab+d-1)\dots(ah-ab-1)}{(d+1)!} \geq \frac{(ah)^{d+1}}{(d+1)!}\left(1-\frac{ab+1}{ah}\right)^{d+1}.$$

Similarly, for $h \geq \left(b + \frac{d}{a}\right)(d+1)$, we have

$$\binom{n_2+d}{d+1} \geq \frac{(ah)^{d+1}}{(d+1)!}\left(1-\frac{2(d+1)(ab+1)}{ah}\right)$$

This proves the lower bound in (5.6). □

Remark 5.1. We will use the following estimate without explicit reference many times in the following proofs.

$$x^\alpha - \log x \geq (1/\alpha)\log(e\alpha), \qquad x, \alpha > 0. \tag{5.8}$$

This can be easily verified by computing the minimum of the function $y \mapsto e^{\alpha y} - y$, $y \in \mathbb{R}$.

5.2. Proof of Theorem 3.1

The proof of Theorem 3.1 requires the following lemma [3, Proposition 1.3].

Lemma 5.3. *Let $d\in\mathbb{N}$, $(Y,\|\cdot\|_Y)$ be a d-dimensional normed linear space, and $B_r=\{x\in Y:\|x\|\le r\}$. Then*

$$d\log(r/(2\epsilon))\le C_{2\epsilon}(B_r,\|\cdot\|_Y)\le H_\epsilon(B,\|\cdot\|_Y)\le d\log\left(\max(3r/\epsilon,1)\right). \tag{5.9}$$

Proof of Theorem 3.1. In this proof, observe

$$\mathfrak{K}_j=\left\{g\in X_j:\ \|g\|\le\Delta_j\right\}.$$

Let $\mathcal{C}_j$ be an $\eta_1=\dfrac{A_{M,p}\epsilon}{2M^{1/p}}$-net for each $\mathfrak{K}_j$, $j=0,\cdots,M-1$. Then it is easily verified that $\prod_{j=0}^{M-1}\mathcal{C}_j$ is an $(A_{M,p}\epsilon/2)$-net for $\prod_{j=1}^{M-1}\mathfrak{K}_j$. Therefore, (3.5) shows that $\mathcal{T}_M^{-1}(\prod_{j=0}^{M-1}\mathcal{C}_j)$ is an ϵ-net for $\mathfrak{K}$ with respect to the norm of X. Since the cardinality of $\mathcal{T}_M^{-1}(\prod_{j=0}^{M-1}\mathcal{C}_j)$ is the same as that of $\prod_{j=0}^{M-1}\mathcal{C}_j$, it follows that

$$H_\epsilon(\mathfrak{K},X)\le\sum_{j=0}^{M}H_{\eta_1}(\mathfrak{K}_j,X).$$

Since each $\mathfrak{K}_j$ is a ball of radius Δ_j in the b_j-dimensional space X_j, Lemma 5.3 leads to

$$H_\epsilon(\mathfrak{K},X)\le\sum_{j=0}^{M}b_j\log\left(\max\left(3\Delta_j/\eta_1,1\right)\right).$$

This proves the second inequality in (3.7).

The proof of the first inequality in (3.7) is similar. We let $\eta_1=2\epsilon/B_{N,r}$ and let $\tilde{\mathcal{C}}_j$ be the maximal η_1-separated subset of each $\mathfrak{K}_j$, $j=0,\cdots,N-1$. Then $\prod_{j=0}^{N-1}\tilde{\mathcal{C}}_j$ is an η_1-separated subset of $\widetilde{\mathfrak{K}}_N$, and hence, (3.5) shows that $\mathcal{T}^{-1}(\prod_{j=0}^{N-1}\tilde{\mathcal{C}}_j)$ an $2\epsilon=\eta_1B_{N,r}$-separated subset of $\mathfrak{K}$. The cardinality of $\mathcal{T}^{-1}(\prod_{j=0}^{N-1}\tilde{\mathcal{C}}_j)$ is the same as that of $\prod_{j=0}^{N-1}\tilde{\mathcal{C}}_j$. Lemma 5.3 then shows that

$$H_\epsilon(\mathfrak{K},X)\ge\sum_{j=0}^{N-1}C_{2\epsilon}(\mathfrak{K}_j,X)\ge\sum_{j=0}^{N-1}b_j\log\left(\frac{\Delta_j}{2B_{N,r}\epsilon}\right).$$

This proves the first equation in (3.7). □

5.3. Proof of Theorem 3.2

In this subsection, we apply Theorem 3.1 with $p=r=2$ to give the proof of Theorem 3.2. In this case, $X=L^2(I^q)$, $X_j=\mathrm{span}\{p_{\mathbf{k}}:\ |\mathbf{k}|_1=j\}$, $\{\Delta_n\}_{n=0}^\infty=\{\rho^n\}_{n=0}^\infty$ and $\mathfrak{K}=\mathcal{A}_\rho$.

Moreover, for each $n\in\mathbb{N}$ and each $f\in\bigoplus_{j=0}^{n-1}X_j$,

$$\begin{aligned}\|\mathcal{T}_n(f)\|_{\Pi,2,n}&=\left|\left(\left\|\mathrm{Proj}_0(f)\right\|_{L^2(I^q)},\ldots,\left\|\mathrm{Proj}_{n-1}(f)\right\|_{L^2(I^q)}\right)\right|_2\\&=\left(\sum_{j=0}^{n-1}\left\|\sum_{|\mathbf{k}|_1=j}\hat{f}(\mathbf{k})p_{\mathbf{k}}\right\|_{L^2(I^q)}^2\right)^{1/2}\\&=\|f\|_{L^2(I^q)}.\end{aligned}$$

Therefore, $A_{n,2}=B_{n,2}=1$.

Proof of Theorem 3.2. In order to apply Theorem 3.1, we need to find an integer larger than $\mathcal{N}(\epsilon/2)$ (cf. (3.8)), which is the solution of the following inequalities:

$$\sum_{n=N}^{\infty}\rho^{2n}=\frac{\rho^{2N}}{1-\rho^2}<\frac{\epsilon^2}{4}\le\frac{\rho^{2N-2}}{1-\rho^2}=\sum_{n=N-1}^{\infty}\rho^{2n}.$$

It gives

$$\mathcal{N}(\epsilon/2) = \left\lfloor \frac{\log\frac{2}{\epsilon} + \log\frac{1}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}} \right\rfloor + 1.$$

Now we estimate the bound in (3.1). For simplicity, write

$$N_1 = \left\lfloor \frac{\log\frac{2}{\epsilon} + \log\frac{1}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}} \right\rfloor. \tag{5.10}$$

Then we can apply the second inequality in (3.1) to $M = N_1 + 1$. For each $n = 0, \ldots, N_1 - 1$,

$$\log\left(\frac{6\Delta_n\sqrt{N_1+1}}{\epsilon}\right) = \log\left(\frac{6\sqrt{N_1+1}\rho^n}{\epsilon}\right) \le \log\left(\frac{6\sqrt{N_1+1}\rho^n}{2\frac{\rho^{N_1+1}}{\sqrt{1-\rho^2}}}\right) \le \log\left(3\rho^{-1}\sqrt{(N_1+1)(1-\rho^2)}\rho^{n-N_1}\right).$$

Hence, by Theorem 3.1 and our condition on ϵ,

$$H_\epsilon\left(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}\right) \le \sum_{n=0}^{N_1}\binom{n+q-1}{q-1}\left(\frac{1}{2}\log(N_1+1) + \log\left(3\rho^{-1}\sqrt{(1-\rho^2)}\right) + (N_1-n)\log\frac{1}{\rho}\right). \tag{5.11}$$

By Lemma 5.1,

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \le \binom{N_1+q}{q+1}\log\frac{1}{\rho} + \binom{N_1+q}{q}\frac{1}{2}\log\left(9(\rho^{-2}-1)(N_1+1)\right). \tag{5.12}$$

Consider the upper bound in (5.12). For $\epsilon < \frac{2\rho}{\sqrt{1-\rho^2}}\left(\frac{2\log\frac{1}{\rho}}{9(\rho^{-2}-1)(q+1)}\right)^{q+1}$, we have

$$N_1 + 1 \ge \frac{\log\frac{2}{\epsilon} + \log\frac{1}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}} \ge \frac{q+1}{\log\frac{1}{\rho}}\log\frac{9(\rho^{-2}-1)(q+1)}{2\log\frac{1}{\rho}}.$$

Taking $\alpha = \frac{N_1}{q+1}\log\frac{1}{\rho}$ and $\alpha' = \log\frac{9(\rho^{-2}-1)(q+1)}{2\log\frac{1}{\rho}}$ in the trivial inequality

$$\alpha \ge \frac{1}{2}(\log(2\alpha) + \alpha'), \quad \alpha \ge \alpha' > 0,$$

we have

$$\frac{N_1}{q+1}\log\frac{1}{\rho} \ge \frac{1}{2}\log\left(\frac{9(\rho^{-2}-1)(q+1)}{\log\frac{1}{\rho}}\frac{N_1+1}{q+1}\log\frac{1}{\rho}\right) = \frac{1}{2}\log\left(9(N_1+1)(\rho^{-2}-1)\right).$$

Hence, (5.12) leads to

$$\begin{aligned}
H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) &\le \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+q)}{q+1}\right)^{q+1}\log\frac{1}{\rho} + \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+q)}{q}\right)^q \times \frac{1}{2}\log\left(9(N_1+1)(1-\rho^2)\right) \\
&\le \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+q)}{q+1}\right)^{q+1}\log\frac{1}{\rho} + \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+q)}{q+1}\right)^q\left(\frac{q+1}{q}\right)^q\frac{N_1+q}{q+1}\log\frac{1}{\rho} \\
&\le \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+q)}{q+1}\right)^{q+1}\log\frac{1}{\rho} + \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+q)}{q+1}\right)^q e\frac{N_1+q}{q+1}\log\frac{1}{\rho} \\
&\le \frac{4e^{q+1}}{\sqrt{2\pi}}\left(\frac{N_1+q}{q+1}\right)^{q+1}\log\frac{1}{\rho}.
\end{aligned}$$

Involving our choice of N_1 (5.10) in this formula,

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \le \frac{4e^{q+1}}{\sqrt{2\pi}}\left(1 + \frac{\log\left(\frac{2\rho}{\sqrt{1-\rho^2}}\frac{1}{\epsilon}\right)}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}$$

Now we prove the asymptotic relation. Applying (5.6) with $h \leftarrow \log\frac{1}{\epsilon}$ to (5.12), and noticing that $\binom{N_1+q}{q} = \binom{N_1+q}{q+1}\frac{q+1}{N_1}$, then for ϵ sufficiently small,

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \le$$
$$\frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}\left(1+\frac{2(q+1)\left(\log\frac{2}{\sqrt{1-\rho^2}}+q\log\frac{1}{\rho}\right)}{\log\frac{1}{\epsilon}}\right)\left(1+\frac{(q+1)\log\left(9(\rho^{-2}-1)(N_1+1)\right)}{2N_1}\right).$$

Bounding $1/N_1$ by $\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}-1\right)^{-1}$ and $\log(N_1+1)$ by $\log\log\frac{1}{\epsilon}$, we get

$$\begin{aligned}
&H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)})\\
&\le\frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}\left(1+\frac{2(q+1)\left(\log\frac{2}{\sqrt{1-\rho^2}}+q\log\frac{1}{\rho}\right)}{\log\frac{1}{\epsilon}}\right)\left(1+(q+1)\frac{\log\frac{1}{\rho}}{\log\frac{1}{\epsilon}}\left(\log\log\frac{1}{\epsilon}+\log(9(\rho^{-2}-1))\right)\right)\\
&\le\frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}\left(1+\frac{2(q+1)\left(\log\frac{2}{\sqrt{1-\rho^2}}+q\log\frac{1}{\rho}\right)}{\log\frac{1}{\epsilon}}+2(q+1)\frac{\log\frac{1}{\rho}}{\log\frac{1}{\epsilon}}\left(\log\log\frac{1}{\epsilon}+\log(9(\rho^{-2}-1))\right)\right)\\
&=\frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}\left[1+2(q+1)\frac{\log\frac{1}{\rho}}{\log\frac{1}{\epsilon}}\left(\left(\frac{\log\frac{2}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}}+q\right)+\left(\log\log\frac{1}{\epsilon}+\log(36\rho^2)-2\log\frac{2}{\sqrt{1-\rho^2}}\right)\right)\right]\\
&<\frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}\left[1+2(q+1)\frac{\log\frac{1}{\rho}}{\log\frac{1}{\epsilon}}\left(\log\log\frac{1}{\epsilon}+\frac{\log\frac{2}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}}+q+4\right)\right].
\end{aligned}$$

Next, we prove the lower bound; for this purpose, we chose N_2 to be as large as we can under the restriction $\log\left(\frac{\delta_{N_2}}{2\epsilon}\right) = \log\left(\frac{\rho^{N_2}}{2\epsilon}\right) \ge 0$.

Solving the inequalities $\rho^{N_2+1} < 2\epsilon \le \rho^{N_2}$, we get

$$N_2 = \left\lfloor \frac{\log\frac{1}{2\epsilon}}{\log\frac{1}{\rho}} \right\rfloor - 1. \tag{5.13}$$

Since $2\epsilon \le \rho^{N_2}$, we have

$$\log\left(\frac{\rho^n}{2\epsilon}\right) \ge \log\left(\frac{\rho^n}{\rho^{N_2}}\right) \ge (N_2-n)\binom{n+q-1}{q-1}\log\frac{1}{\rho}, \quad n \le N_2.$$

Now by Theorem 3.1 and (5.3),

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \ge \binom{N_2+q}{q+1}\log\frac{1}{\rho}. \tag{5.14}$$

Similarly as before, involving $N_2 = \left\lfloor \frac{\log\frac{1}{2\epsilon}}{\log\frac{1}{\rho}} \right\rfloor$ in (5.14), using (5.4),

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \ge \frac{2^{q+1}}{8\sqrt{2\pi(q+1)}}\left(1+\frac{\log\frac{\rho^2}{2\epsilon}}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}.$$

This completes the proof of (3.11).

Consider the asymptotic relation. For ϵ sufficiently small,

$$H_\epsilon(\mathcal{A}_\rho, \|\cdot\|_{L_2(I^q)}) \ge \binom{N_2+q}{q+1}\log\frac{1}{\rho} \ge \frac{\log\frac{1}{\rho}}{(q+1)!}\left(\frac{\log\frac{1}{\epsilon}}{\log\frac{1}{\rho}}\right)^{q+1}\left(1-\frac{2(q+1)\log\frac{2}{\rho}}{\log\frac{1}{\epsilon}}\right).$$

The two inequalities prove (3.13). □

5.4. Proof of Theorem 3.3: upper bound

As in the previous subsection, we apply Theorem 3.1 to give the proof. We recall the condition (3.17) relating Q, τ, and ϵ.

Proof. For the upper bound, we apply Theorem 3.1 with $p=1$, $X=L_\infty(I_{\mathbf r})$, $X_j=\mathrm{span}\{p_{\mathbf{k},\mathbf{r}}:\ |\mathbf{k}|_1=j\}$,

$$\Delta_N=\Lambda(N)=CN^{Q/2}\frac{\tau^N}{N!},\quad N=0,1,\dots$$

and $\mathfrak{K}=\mathcal{B}_Q$.

In this case, for $F\in\bigoplus_{j=0}^{n-1}X_j$,

$$\|\mathcal{T}_n(F)\|_{\Pi,1,n}=\sum_{k=0}^{n-1}\left\|\mathrm{Proj}_k(F)\right\|_{L_\infty(I_{\mathbf r})}\geq\|F\|_{L_\infty(I_{\mathbf r})},$$

which means we can take $A_{n,1}=1$ for each $n\in\mathbb{N}$. Now we only need to find a proper $M\geq\mathcal{N}(\epsilon/2)$ to apply Theorem 3.1.

To estimate the sum of $\Lambda(N)$, we investigate the decay of this sequence.

$$\frac{\Lambda(N+1)}{\Lambda(N)}=\left(\frac{N+1}{N}\right)^{Q/2}(N+1)^{-1}\tau=\frac{\tau}{N+1}\left(\left(1+\frac{1}{N}\right)^N\right)^{\frac{Q}{2N}}\leq\frac{\tau}{N+1}\exp\left(\frac{Q}{2N}\right).$$

Then for $N\geq Q/\left(\log\left(\frac{Q}{2\tau}\right)\right)$,

$$\frac{\Lambda(N+1)}{\Lambda(N)}\leq\frac{\tau\log\left(\frac{Q}{2\tau}\right)}{Q}\exp\left(\frac{\log\left(\frac{Q}{2\tau}\right)}{2}\right)\leq\frac{\tau}{Q}\log\left(\frac{Q}{2\tau}\right)\sqrt{\frac{Q}{2\tau}}\leq\frac{\tau}{Q}\left(\sqrt{\frac{Q}{2\tau}}\right)^2=\frac{1}{2}$$

and consequently

$$\sum_{N=M+1}^{\infty}\Lambda(N)\leq2\Lambda(M+1),\quad M\geq Q/\left(\log\left(\frac{Q}{2\tau}\right)\right).\tag{5.15}$$

This enables us to find a proper $N_1\geq\mathcal{N}(\epsilon/2)$. To do this, we only need to find a proper integer N_1 with $2\Lambda(N_1+1)\leq\epsilon/2$. Applying Stirling's estimation, we have for any $M>\frac{Q-1}{2}$ and $M_0=M-\frac{Q-1}{2}$,

$$\begin{aligned}2\Lambda(M+1)\leq&2C(M+1)^{Q/2}\tau^{M+1}\frac{1}{\sqrt{2\pi(M+1)}}\left(\frac{e}{M+1}\right)^{M+1}\\ \leq&2C(M+1)^{\frac{Q-1}{2}}\left(\frac{e\tau}{M+1}\right)^{M+1}\leq2C\left(\frac{e\tau}{M_0+1}\right)^{M_0+1}(e\tau)^{\frac{Q-1}{2}}.\end{aligned}$$

So it suffices to find M_0 such that

$$2C\left(\frac{e\tau}{M_0+1}\right)^{M_0+1}(e\tau)^{\frac{Q-1}{2}}\leq\epsilon/2.$$

This inequality is equivalent to the inequality

$$\frac{M_0+1}{e\tau}\log\left(\frac{M_0+1}{e\tau}\right)\geq\frac{\log\frac{4}{\epsilon}+\log\left(C(e\tau)^{\frac{Q-1}{2}}\right)}{e\tau}.$$

Under the condition that

$$\epsilon\leq\left(\frac{2\pi e\tau}{Q}\right)^{Q/2}\frac{4}{(e\tau)^{1/2}\exp(e^2\tau)},$$

we have

$$\frac{M_0+1}{e\tau}\geq2\frac{\left(\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)(e\tau)^{-1}}{\log\left(\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}$$

for any

$$M_0 \geq 2\frac{\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}-1. \tag{5.16}$$

Note $x \geq \frac{2y}{\log y} \Rightarrow x\log x \geq y$ for all $y > e$, we conclude

$$\frac{M_0+1}{e\tau}\log\left(\frac{M_0+1}{e\tau}\right) \geq \frac{\log\frac{4}{\epsilon}+\frac{Q}{2}\log\left(\frac{2e\pi\tau}{Q}\right)}{e\tau}$$

holds true for

$$\epsilon < \left(\frac{2\pi e\tau}{Q}\right)^{Q/2}\frac{4}{(e\tau)^{1/2}\exp(e^2\tau)}.$$

Then $2\Lambda\left(M_0+\frac{Q-1}{2}+1\right) \leq \epsilon/2$ for M_0 satisfying (5.16).

Therefore, in order to make $\sum_{N=N_1+1}^{\infty}\Lambda(N) \leq \epsilon/2$ hold true, it suffices to take

$$N_1 = \left\lfloor 2\frac{\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\epsilon}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}+\frac{Q-1}{2}\right\rfloor, \tag{5.17}$$

then N_1+1 is a proper integer for which Theorem 3.1 can be applied. Now we use Stirling's approximation to bound $\Lambda(N)$ by

$$CN^{Q/2}\frac{\tau^N}{N!} \leq C(e\tau)^N\left(\frac{1}{N}\right)^{N-Q/2} \leq C(e\tau)^N(N_1+1)^{Q/2},\ N \leq N_1+1$$

and notice that

$$\begin{aligned}\frac{6C(e\tau)^N(N_1+1)^{Q/2}(N_1+1)}{\epsilon} &\geq 6(e\tau)^N\left(\frac{2\pi}{Q}\right)^{Q/2}(Q/2)^{Q/2+1}\left(\frac{Q}{2\pi e\tau}\right)^{Q/2}\frac{(e\tau)^{1/2}\exp(e^2\tau)}{4}\\ &\geq 6(e\tau)^N\pi^{Q/2}e^{Q/4}\frac{(e\tau)^{1/2}\exp(e^2\tau)}{4} > 1.\end{aligned}$$

Therefore,

$$\begin{aligned}&\log\left(\max\left\{\frac{6\Lambda(N)(N_1+1)}{\epsilon},1\right\}\right) \leq \log\left(\frac{6C(e\tau)^N(N_1+1)^{Q/2+1}}{\epsilon}\right)\\ &=\left(\log\frac{6C}{\epsilon}+N\log(e\tau)+\frac{Q+2}{2}\log(N_1+1)\right)\end{aligned}$$

and we can apply Lemma 5.1 to get

$$\begin{aligned}H_\epsilon\left(\mathcal{B}_Q,\|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) &\leq \sum_{N=0}^{N_1}\binom{N+Q-1}{Q-1}\left(\log\frac{6C}{\epsilon}+N\log(e\tau)+\frac{Q+2}{2}\log(N_1+1)\right)\\ &\leq Q\binom{N_1+Q}{Q+1}\log(e\tau)+\binom{N_1+Q}{Q}\left(\log\frac{6C}{\epsilon}+\frac{Q+2}{2}\log(N_1+1)\right).\end{aligned} \tag{5.18}$$

Observing

$$\binom{N_1+Q}{Q+1}Q < N_1\binom{N_1+Q}{Q},$$

$$H_\epsilon\left(\mathcal{B}_Q,\|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) \leq \binom{N_1+Q}{Q}\left(N_1\log(e\tau)+\log\frac{6C}{\epsilon}+\frac{Q+2}{2}\log(N_1+1)\right). \tag{5.19}$$

Next, we express the bound (5.19) in terms of ϵ. We will apply Lemma 5.2.

In this proof only, let

$$B = \log\left(\frac{4}{\epsilon}\left(\frac{2e\pi\tau}{Q}\right)^{\frac{Q}{2}}\right),$$

then

$$N_1 = \left\lfloor \frac{2B}{\log B - \log(e\tau)} + \frac{Q-1}{2} \right\rfloor \geq 2e^2 + \frac{Q-1}{2}.$$

Now we can apply (5.4) to conclude

$$\begin{aligned}
H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) &\leq \frac{2}{\sqrt{2\pi}}\left(\frac{e(N_1+Q)}{Q}\right)^Q\left(N_1\log(e\tau) + \log\frac{6C}{\epsilon} + \frac{Q+2}{2}\log(N_1+1)\right)\\
&\leq \frac{2}{\sqrt{2\pi}}\frac{e^Q 2^Q}{Q^Q}\left(\frac{B}{\log B - \log(e\tau)} + \frac{3Q}{4}\right)^Q\\
&\times\left[\left(2\frac{B}{\log B-\log(e\tau)} + \frac{Q-1}{2}\right)\log(e\tau) + B + 1 - \frac{Q}{2}\log(e\tau) + \frac{Q+2}{2}\log\left(\frac{2B}{\log B - \log(e\tau)} + \frac{Q+1}{2}\right)\right].
\end{aligned}$$

Bounding

$$B + 1 - \frac{Q}{2}\log(e\tau) \leq \left(\frac{2B}{\log B - \log e\tau} + \frac{3Q}{2}\right)\frac{\log B - \log e\tau}{2}$$

and

$$\begin{aligned}
\frac{Q+2}{2}\log\left(\frac{2B}{\log B - \log(e\tau)} + \frac{Q+1}{2}\right) &\leq \left(3 + \frac{3Q}{2}\right)\times\frac{1}{3}\left(\log(2B) + \log\frac{Q+1}{2}\right)\\
&\leq \left(\frac{2B}{\log B - \log(e\tau)} + \frac{3Q}{2}\right)\times\frac{1}{3}\left(\log(2B) + \log\frac{Q+1}{2}\right),
\end{aligned}$$

we have

$$\begin{aligned}
H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) &\leq \frac{2}{\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{B}{\log B - \log(e\tau)} + \frac{3Q}{4}\right)^Q\\
&\times\left(\frac{2B}{\log B - \log(e\tau)} + \frac{3Q}{2}\right)\left[\log(e\tau) + \frac{\log B - \log(e\tau)}{2} + \frac{1}{3}\left(\log 2B + \log\frac{Q+1}{2}\right)\right]\\
&\leq \frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{\log\frac{4}{\epsilon} + \frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\epsilon} + \frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right) - \log(e\tau)} + \frac{3Q}{4}\right)^{Q+1}\left(7\log\log\frac{1}{\epsilon} + 2\log(Q+1) + 6\log(e\tau)\right).
\end{aligned}$$

Finally, consider the asymptotic relation of the bound when $\epsilon \to 0$. A simple observation shows

$$N_1 = 2\frac{\log\frac{1}{\epsilon}}{\log\log\frac{1}{\epsilon}}(1 + o(1))$$

as $\epsilon \to 0$.

Then (5.19) gives

$$\begin{aligned}
H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) &\leq \frac{(N_1+Q)^Q}{Q!}\left(2\log\frac{1}{\epsilon}\right)(1+o(1))\\
&\leq \frac{1}{Q!}\left((1+o(1))\frac{2\log\frac{1}{\epsilon}}{\log\log\frac{1}{\epsilon}}\right)^Q\left(2\log\frac{1}{\epsilon}\right)(1+o(1))\\
&\leq \frac{1}{Q!}\frac{\left(2\log\frac{1}{\epsilon}\right)^{Q+1}}{\left(\log\log\frac{1}{\epsilon}\right)^Q}(1+o(1)). \quad \square
\end{aligned}$$

5.5. Proof of Theorem 3.3: lower bound

In this section, we consider the lower bound. As in the last subsection, we apply Theorem 3.1 with $p=1$, $X=L_\infty(I_{\mathbf{r}})$, $X_j=\mathrm{span}\{p_{\mathbf{k},\mathbf{r}}: \ |\mathbf{k}|_1=j\}$,

$$\Delta_N=\Lambda(N)=CN^{Q/2}\frac{\tau^N}{N!}, \quad N=0,1,\dots$$

and $\mathfrak{K}=\mathcal{B}_Q$. We recall also the condition (3.19).

Proof. For Chebyshev polynomials, by [6, Section 12, Chapter 2], we have

$$\|S_n(F)\|_{L_\infty(I_{\mathbf{r}})}\le(\log n+1)^Q\|F\|_{L_\infty(I_{\mathbf{r}})}, \quad n\in\mathbb{N}.$$

Then for $F\in\bigoplus_{j=0}^{n-1}X_j$,

$$\|\mathcal{T}_n(F)\|_{\Pi,1,n}=\sum_{k=0}^{n-1}\left\|\mathrm{Proj}_k(F)\right\|_{L_\infty(I_{\mathbf{r}})}\le n(\log n+1)^Q\|F\|_{L_\infty(I_{\mathbf{r}})}, \quad n\ge 1,$$

which means we can take $B_{n,1}=n(\log n+1)^Q$ for each $n\in\mathbb{N}$. Now we only need to find a proper N_2 to apply Theorem 3.1.

Like in the proof of Theorem 3.2, our principle of choosing N_2 is finding it as large as we can under the restriction

$$\log\left(\frac{\Lambda(N_2)}{2(N_2+1)(\log N_2+1)^Q\epsilon}\right)\ge 0.$$

To find a solution of

$$\frac{\Lambda(N)}{2(N+1)(\log N+1)^Q}\ge\epsilon,$$

we make the Stirling's estimation

$$\Lambda(N)\ge CN^{Q/2}\tau^N\frac{1}{2\sqrt{2\pi N}}\left(\frac{e}{N}\right)^N=\frac{C(e\tau)^N}{2\sqrt{2\pi}}\left(\frac{1}{N}\right)^{N-\frac{Q-1}{2}}:=\Lambda_0(N).$$

Then

$$\begin{aligned}\frac{\Lambda(N)}{2(N+1)(\log N)^Q}&\ge\frac{\Lambda_0(N)}{2(N+1)(\log N)^Q}=\frac{C(e\tau)^{-\frac12}}{2\sqrt{2\pi}}\left(\frac{e\tau}{N}\right)^{N+\frac12}\left(\frac{N}{(\log N)^2}\right)^{Q/2}\frac{1}{N+1}\\&\ge\frac{C}{4\sqrt{2\pi e\tau}}\left(\frac{e\tau}{N}\right)^{N+\frac12}.\end{aligned}$$

In this proof only, let

$$B=\log\left(\frac{C}{4\sqrt{2\pi e\tau}\epsilon}\right),$$

then it suffices to find a solution of

$$e^B\left(\frac{e\tau}{N}\right)^{N+\frac12}\ge 1.$$

Taking logarithms on both sides, we conclude it suffices to solve

$$\left(N+\frac12\right)\log\frac{N}{e\tau}\le B. \tag{5.20}$$

Let

$$N_2=\left\lfloor\frac{B}{\log B-\log(e\tau)}-\frac12\right\rfloor. \tag{5.21}$$

Note $x\le\frac{y}{\log y}\Rightarrow x\log x\le y$ for all $y\ge e$, and it is clear that $\frac{B}{e\tau}\ge e$ under the condition that

$$\epsilon\le\left(\frac{2\pi}{Q}\right)^{Q/2}\frac{1}{4\sqrt{2\pi e\tau}}\xi_\tau^{-2\xi_\tau}\le\frac{1}{4\sqrt{2\pi e\tau}}\exp(-e^2\tau)\left(\frac{2\pi}{Q}\right)^{Q/2},$$

then

$$\frac{N_2+\frac{1}{2}}{e\tau}\log\frac{N_2+\frac{1}{2}}{e\tau}\le\frac{B}{e\tau}.$$

Consequently, N_2 is a solution of (5.20), hence, a solution of

$$\frac{\Lambda(N)}{2(N+1)(\log N)^Q}\ge\epsilon.$$

Now

$$\frac{\Lambda(N)}{2\epsilon(N_2+1)\log N_2}\ge\frac{\Lambda_0(N)}{2\epsilon(N_2+1)(\log N_2+1)^Q}=e^B\left(\frac{e\tau}{N}\right)^{N-\frac{Q-1}{2}}(\log N_2+1)^{-Q}.$$

Together with Theorem 3.1,

$$\begin{aligned}H_\epsilon\left(\mathcal{B}_Q,\|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right)&\ge\sum_{N=0}^{N_2}\binom{N+Q-1}{Q-1}\log\left(e^B\left(\frac{e\tau}{N}\right)^{N-\frac{Q-1}{2}}(\log N_2+1)^{-Q}\right)\\&=\sum_{N=0}^{N_2}\binom{N+Q-1}{Q-1}\left[\log\left(e^B\left(\frac{e\tau}{N}\right)^{N+\frac{1}{2}}\right)+\frac{Q}{2}\log\frac{N}{e\tau}-Q\log(\log N_2+1)\right].\end{aligned}\tag{5.22}$$

On one hand,

$$\begin{aligned}&\sum_{N=0}^{N_2}\binom{N+Q-1}{Q-1}\left[\frac{Q}{2}\log\frac{N}{e\tau}-Q\log(\log N_2+1)\right]\\&\qquad\ge\sum_{n=\lfloor N_2/2\rfloor}^{N_2}\binom{N+Q-1}{Q-1}\frac{Q}{2}\log\frac{N}{e\tau}-\binom{N_2+Q}{Q}Q\log(\log N_2+1)\\&\qquad\ge\sum_{n=\lfloor N_2/2\rfloor}^{N_2}\binom{N+Q-1}{Q-1}\frac{Q}{2}\log\frac{N_2}{3e\tau}-\binom{N_2+Q}{Q}Q\log(\log N_2+1)\\&\qquad\ge\frac{1}{2}\binom{N_2+Q}{Q}\frac{Q}{2}\log\frac{N_2}{3e\tau}-\binom{N_2+Q}{Q}Q\log(\log N_2+1)\\&\qquad=\frac{Q}{4}\binom{N_2+Q}{Q}\log\frac{N_2}{3e\tau(\log N_2+1)^4}.\end{aligned}$$

In this proof only, let $a_\tau=\max\{128,3e^2\tau\}$. Since $\epsilon<\left(\frac{2\pi}{Q}\right)^{Q/2}\frac{1}{4\sqrt{2\pi e\tau}}\xi_\tau^{-2\xi_\tau}$, we have

$$B\ge 2\frac{16a_\tau(\log a_\tau)^4+2e}{e}\log\frac{16a_\tau(\log a_\tau)^4+2e}{e}.$$

For $y\ge e$, we have $x\ge 2y\log y\Rightarrow\frac{x}{\log x}\ge y$, so

$$\frac{B}{\log B-\log(e\tau)}\ge\frac{B}{\log B}\ge\frac{16a_\tau(\log a_\tau)^4+2e}{e}.$$

Therefore,

$$N_2\ge\frac{16a_\tau(\log a_\tau)^4}{e}.$$

We have also $x\ge 16y(\log y)^4\Rightarrow\frac{x}{(\log x)^4}\ge y$ for $y\ge 128$, then

$$\frac{eN_2}{\log(eN_2)^4}\ge a_\tau\ge 3e^2\tau,$$

this is

$$\frac{N_2}{3e\tau(\log N_2+1)^4}\ge 1.$$

Consequently,

$$\sum_{N=0}^{N_2} \binom{N+Q-1}{Q-1} \left[\frac{Q}{2} \log \frac{N}{e\tau} - Q \log\left(\log N_2 + 1\right) \right] \geq 0.$$

On the other hand, since $e^B \left(\frac{e\tau}{N_2}\right)^{N_2+\frac{1}{2}} \geq 1$, we have for $N \leq \lfloor N_2/2 \rfloor$,

$$e^B \left(\frac{e\tau}{N}\right)^{N+\frac{1}{2}} \geq \left(\frac{e\tau}{N}\right)^{N+\frac{1}{2}} \left(\frac{N_2}{e\tau}\right)^{N_2+\frac{1}{2}} = \left(\frac{N_2}{N}\right)^{N+\frac{1}{2}} \left(\frac{N_2}{e\tau}\right)^{N_2-N} \geq \left(\frac{N_2}{e\tau}\right)^{\frac{N_2}{2}}.$$

Then

$$H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) \geq \sum_{N=0}^{\lfloor N_2/2 \rfloor} \binom{N+Q-1}{Q-1} \frac{N_2}{2}(\log N_2 - \log e\tau) + \frac{Q}{4}\binom{N_2+Q}{Q} \log \frac{N_2}{3e\tau\left(\log N_2+1\right)^4}.$$

By Lemma 5.1,

$$H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) \geq \frac{1}{2}\binom{\lfloor N_2/2 \rfloor + Q}{Q} N_2(\log N_2 - \log e\tau) + \frac{Q}{4}\binom{N_2+Q}{Q} \log \frac{N_2}{3e\tau\left(\log N_2+1\right)^4}. \tag{5.23}$$

Next we express the bound (5.23) in terms of ϵ. To begin with, we see that

$$\left\lfloor \frac{N_2}{2} \right\rfloor \geq \frac{1}{2}N_2 - \frac{1}{2} \geq \frac{1}{2}\left(\frac{B}{\log B - \log(e\tau)} - \frac{1}{2} - 1\right) - \frac{1}{2} = \frac{1}{2}\frac{B}{\log B - \log(e\tau)} - \frac{5}{4}.$$

Apply Lemma 5.2 and substitute (5.21),

$$\begin{aligned}
&H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) \geq \frac{1}{2}\frac{1}{8\sqrt{\pi Q}} \left(\frac{2(\lfloor N_2/2 \rfloor + Q)}{Q}\right)^Q N_2(\log N_2 - \log e\tau) \\
&\geq \frac{1}{16\sqrt{\pi Q}} \left(\frac{2}{Q}\right)^Q \left(\frac{1}{2}\frac{B}{\log B - \log(e\tau)} - \frac{5}{4} + Q\right)^Q \left(\frac{B}{\log B - \log(e\tau)} - \frac{3}{2} - \log(e\tau)\right) \\
&\geq \frac{1}{16\sqrt{\pi Q}} \left(\frac{1}{Q}\right)^Q \left(\frac{\log \frac{1}{4\sqrt{2\pi e\tau\epsilon}} + \frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)}{\log\left(\log \frac{1}{4\sqrt{2\pi e\tau\epsilon}} + \frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)\right) - \log(e\tau)} - \frac{5}{2} + 2Q\right)^Q \\
&\times \left(\frac{\log \frac{1}{4\sqrt{2\pi e\tau\epsilon}} + \frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)}{\log\left(\log \frac{1}{4\sqrt{2\pi e\tau\epsilon}} + \frac{Q}{2}\log\left(\frac{2\pi}{Q}\right)\right) - \log(e\tau)} - \frac{3}{2} - \log(e\tau)\right).
\end{aligned}$$

For the asymptotic relation of the bound when $\epsilon \to 0$, a simple observation shows

$$N_2 = \frac{\log \frac{1}{\epsilon}}{\log\log \frac{1}{\epsilon}}(1+o(1))$$

as $\epsilon \to 0$. Similarly as before, (5.23) gives

$$\begin{aligned}
H_\epsilon\left(\mathcal{B}_Q, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}\right) &\geq \frac{1}{2}\frac{(\lfloor N_2/2 \rfloor)^Q}{Q!} \frac{\log \frac{1}{\epsilon}}{\log\log \frac{1}{\epsilon}} \left(\log \frac{\log \frac{1}{\epsilon}}{\log\log \frac{1}{\epsilon}}\right)(1+o(1)) \\
&\geq \frac{1}{2Q!}\left(\frac{\log \frac{1}{\epsilon}}{2\log\log \frac{1}{\epsilon}}\right)^Q \left(\frac{\log \frac{1}{\epsilon}}{\log\log \frac{1}{\epsilon}}\right)\left(\log\log \frac{1}{\epsilon} - \log\log\log \frac{1}{\epsilon}\right)(1+o(1)) \\
&\geq \frac{1}{2Q!}\left(\frac{\log \frac{1}{\epsilon}}{2\log\log \frac{1}{\epsilon}}\right)^Q \left(\log \frac{1}{\epsilon}\right)(1+o(1)).
\end{aligned}$$

This proves (3.21). □

5.6. Proof of Theorem 3.4

Proof. Since $\epsilon < \frac{\sqrt{1-\rho^2}}{4}\rho^q$, by Theorem 3.2, the $\epsilon/2$-entropy of $\mathcal{A}_\rho$ can be bounded by

$$H_{\epsilon/2}(\mathcal{A}_\rho, \|\cdot\|_{L^2(I^q)}) \le \frac{4e^{q+1}}{\sqrt{2\pi}}\left(1+\frac{\log\left(\frac{2\rho}{\sqrt{1-\rho^2}}\frac{2}{\epsilon}\right)}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}\log\frac{1}{\rho} \le \frac{4(2e)^{q+1}}{\sqrt{2\pi}}\left(\frac{\log\frac{1}{\epsilon}}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}\log\frac{1}{\rho}. \tag{5.24}$$

Consider the upper bound of the $\epsilon/2$-entropy of $\mathcal{F}$. By taking n as the integer N_1 in the proof of Theorem 3.2,

$$n = \left\lfloor \frac{\log\frac{1}{\epsilon}+\log\frac{4}{\sqrt{1-\rho^2}}}{\log\frac{1}{\rho}} \right\rfloor,$$

we get from there that $\|f - s_{n+1}(f)\|_{L^2(I^q)} \le \epsilon/4$ holds for all $f \in \mathcal{A}_\rho$. Now n is fixed in the rest of the proof. For convenience, denote $\tau = \sum_{j=1}^{Q} v_j r_j$. Then $\tau = \frac{Q}{2e^{3/2}\pi}$.

In this case,

$$\left\|\tilde{F} - \tilde{F}\circ s_{n+1}\right\| = \sup_{f\in\mathcal{A}_\rho}\left\|\tilde{F}(f) - \tilde{F}(s_{n+1}(f))\right\| \le \sup_{f\in\mathcal{A}_\rho}\|f - s_{n+1}(f)\|_{L^2(I^q)} \le \epsilon/4.$$

Thus, any $\epsilon/4$-cover of the set $\{F\circ s_{n+1} : \tilde{F}\in\mathcal{F}\}\subset\mathcal{F}$ is an $\epsilon/2$-cover of $\mathcal{F}$.

The map $\tilde{F}\mapsto F$ denoted by

$$\tilde{F}(f) = F\left(\left(\hat{f}(\mathbf{k})\right)_{|\mathbf{k}|_1\le n}\right), \quad f\in\mathcal{A}_\rho$$

is an isometry from $\{F\circ s_{n+1} : \tilde{F}\in\mathcal{F}\}\subset\mathcal{F}$ to $\tilde{\mathcal{B}}_n$ with the $L_\infty(I_{\mathbf{r}})$ norm. Therefore, for the entropy of the former, we only need to consider the entropy of $\tilde{\mathcal{B}}_n$.

Therefore, the ϵ-entropy of $\mathcal{F}\times\mathcal{A}_\rho$ is bounded by

$$H_\epsilon(\mathcal{F}\times\mathcal{A}_\rho, \|\cdot\|) \le H_{\epsilon/2}(\mathcal{A}_\rho, \|\cdot\|_{L^2(I^q)}) + H_{\epsilon/4}(\tilde{\mathcal{B}}_n, \|\cdot\|_{L_\infty(I_{\mathbf{r}})}). \tag{5.25}$$

Let

$$\eta = \min\left\{\frac{\epsilon}{4}, \left(\frac{2\pi e\tau}{Q}\right)^{Q/2}\frac{4}{(e\tau)^{1/2}\exp(e^2\tau)}\right\}.$$

Then using Theorem 3.2 and Theorem 3.3, we conclude

$$\begin{aligned}H_\epsilon(\mathcal{F}\times\mathcal{A}_\rho, \|\cdot\|) \le{}& \frac{4e^{q+1}}{\sqrt{2\pi}}\left(1+\frac{\log\left(\frac{2\rho}{\sqrt{1-\rho^2}}\frac{2}{\epsilon}\right)}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}\log\frac{1}{\rho}\\ &+\frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}+\frac{3Q}{4}\right)^{Q+1}\left(7\log\log\frac{1}{\eta}+\log\left((Q+1)^2(e\tau)^6\right)\right)\end{aligned} \tag{5.26}$$

with $Q = \binom{n+q}{q}$.

Substituting $\tau = \frac{Q}{2e^{3/2}\pi}$ into η and noticing that $Q \gg \log\frac{1}{\epsilon}$,

$$\eta = \min\left\{\frac{\epsilon}{4}, e^{Q/4}4\left(\frac{2\sqrt{e}\pi}{Q}\right)^{1/2}\exp\left\{-\frac{\sqrt{e}Q}{2\pi}\right\}\right\} = 4\left(\frac{2\sqrt{e}\pi}{Q}\right)^{1/2}\exp\left\{-Q\left(\frac{\sqrt{e}}{2\pi}-\frac{1}{4}\right)\right\}.$$

Then

$$\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}=\log\left(\left(\frac{Q}{2\sqrt{e\pi}}\right)^{1/2}\exp\left\{Q\left(\frac{\sqrt{e}}{2\pi}-\frac{1}{4}\right)\right\}\left(\frac{2e\pi}{Q}\frac{Q}{2e^{3/2}\pi}\right)^{Q/2}\right)$$
$$=\log\left(\left(\frac{Q}{2\sqrt{e\pi}}\right)^{1/2}\exp\left\{Q\frac{\sqrt{e}}{2\pi}\right\}\right)=e^2\tau+\frac{1}{2}\log(e\tau).$$

Substituting

$$\tau=\frac{Q}{2e^{3/2}\pi}$$

and

$$2\log(Q+1)+6\log(e\tau)=\log\left((Q+1)^2\left(\frac{Q}{2e^{3/2}\pi}\right)^6\right)\le\log(Q^8)$$

into the latter term in (5.26), we can bound this term as

$$\begin{aligned}&\frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}+\frac{3Q}{4}\right)^{Q+1}\left(7\log\log\frac{1}{\eta}+\log\left((Q+1)^2(e\tau)^6\right)\right)\\
\le&\frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(e^2\tau+\frac{1}{2}\log(e\tau)+\frac{3}{4}Q\right)^{Q+1}(7\log(Q\log Q)+2\log(Q+1)+6\log(e\tau))\\
\le&\frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(e^2\frac{Q}{2e^{3/2}\pi}+\frac{1}{2}\log\left(e\frac{Q}{2e^{3/2}\pi}\right)+\frac{3}{4}Q\right)^{Q+1}\left(7\log(Q\log Q)+\log(Q^8)\right)\\
\le&\frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{\sqrt{e}Q}{2\pi}+\frac{Q}{4}+\frac{3}{4}Q\right)^{Q+1}\left(7\log(Q\log Q)+\log(Q^8)\right)\\
\le&\frac{2}{3\sqrt{2\pi}}\left(2e\left(\frac{\sqrt{e}}{2\pi}+1\right)\right)^Q\left(\frac{\sqrt{e}}{2\pi}+1\right)Q\times 15\log(Q\log Q)\\
\le&\frac{30}{3\sqrt{2\pi}}\left(2e\left(\frac{\sqrt{e}}{2\pi}+1\right)\right)^Q\left(\frac{\sqrt{e}}{2\pi}+1\right)Q\log(Q\log Q).\end{aligned}$$

The fact that $\epsilon<\frac{\sqrt{1-\rho^2}}{4}\rho^q$ implies

$$\log\frac{1}{\epsilon}+\log\frac{4}{\sqrt{1-\rho^2}}+q\log\frac{1}{\rho}\le 2\log\frac{1}{\epsilon}.$$

Then we can bound Q by (5.4) and get

$$Q\le\frac{2}{\sqrt{2\pi}}\left(\frac{e(L+q)}{q}\right)^q\le e^q\left(\frac{\log\frac{1}{\epsilon}+\log\frac{4}{\sqrt{1-\rho^2}}}{q\log\frac{1}{\rho}}+1\right)^q\le\left(\frac{2e\log\frac{1}{\epsilon}}{q\log\frac{1}{\rho}}\right)^q.$$

Consequently,

$$\begin{aligned}&\frac{2}{3\sqrt{2\pi}}\left(\frac{2e}{Q}\right)^Q\left(\frac{\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}}{\log\left(\log\frac{4}{\eta}+\frac{Q}{2}\log\frac{2e\pi\tau}{Q}\right)-\log(e\tau)}+\frac{3Q}{4}\right)^{Q+1}\left(7\log\log\frac{1}{\eta}+\log\left((Q+1)^2(e\tau)^6\right)\right)\\
\le&\frac{60q}{3\sqrt{2\pi}}\exp\left\{\gamma^q\log\left(\frac{e^{3/2}}{\pi}+2e\right)\right\}\gamma^q\log(q\gamma\log\gamma).\end{aligned}$$

Combining this with (5.24) and substituting the values of C and τ into the inequality,

$$
\begin{aligned}
&H_\epsilon(\mathcal{F}\times\mathcal{A}_\rho, \|\cdot\|) \le \frac{4(2e)^{q+1}}{\sqrt{2\pi}}\left(\frac{\log\frac{1}{\epsilon}}{(q+1)\log\frac{1}{\rho}}\right)^{q+1}\log\frac{1}{\rho}\\
&\quad+\frac{60q}{3\sqrt{2\pi}}\exp\left\{\gamma^q\log\left(\frac{e^{3/2}}{\pi}+2e\right)\right\}\gamma^q\log(q\gamma\log\gamma)\\
&\le\frac{4}{\sqrt{2\pi}}\gamma^{q+1}\log\frac{1}{\rho}\left(\frac{q}{q+1}\right)^{q+1}+\frac{60q}{3\sqrt{2\pi}}\exp\left\{\gamma^q\log\left(\frac{e^{3/2}}{\pi}+2e\right)\right\}\gamma^q\log(q\gamma\log\gamma)\\
&\le\gamma^{q+1}\left(\frac{q}{q+1}\right)^{q+1}\log\frac{1}{\rho}+\frac{60q}{3\sqrt{2\pi}}\exp\left\{\gamma^q\log\left(\frac{e^{3/2}}{\pi}+2e\right)\right\}\gamma^q\log(q\gamma\log\gamma),
\end{aligned}
$$

where

$$\gamma=\frac{2e\log\frac{1}{\epsilon}}{q\log\frac{1}{\rho}}.$$

This completes the proof. □

6. Conclusions

We studied the question of which algorithms and data sets are close to each other in terms of some performance metrics. The problem was formulated mathematically rigorously as finding an optimal ϵ-net for a tensor product of two (infinite-dimensional) sets: one representing the data sets and one the algorithms. We solved this problem under certain simplifying assumptions.

Appendix A. Degree of approximation

A.1. Analytic functions

The following lemma is a straightforward consequence of the corresponding well-known one-dimensional results.

Lemma A.1. (a) *If* $\boldsymbol{r}>0$, $\mathbf{k}\in\mathbb{N}^d$ *then for* $\mathbf{z}\in\mathbb{C}^d\setminus I_{\boldsymbol{r}}$, *we have*

$$|p_{\mathbf{k},\mathbf{r}}(\mathbf{z})|\le(\boldsymbol{r})^{-\mathbf{k}}\prod_{j=1}^{d}\left|z_j+\sqrt{z_j^2-r_j^2}\right|^{k_j}. \tag{A.1}$$

(b) *If* $0<\boldsymbol{\rho}<1$, f *is analytic on the closure of* $U_{\boldsymbol{\rho}}$, $\partial\Gamma_{\boldsymbol{\rho}}$ *is the boundary of* $\Gamma_{\boldsymbol{\rho}}$ *and* $g(\mathbf{w})=f((\mathbf{w}+\mathbf{w}^{-1})/2)$, *then*

$$\hat{f}(\mathbf{k})=\frac{1}{(2\pi i)^d}\int\limits_{\partial\Gamma_{\boldsymbol{\rho}}}\frac{g(\mathbf{w})}{\mathbf{w}^{\mathbf{k}+1}}d\mathbf{w}. \tag{A.2}$$

In particular,

$$\left|\hat{f}(\mathbf{k})\right|\le(\boldsymbol{\rho})^{\mathbf{k}}\max_{\mathbf{w}\in\partial\Gamma_{\boldsymbol{\rho}}}|g(\mathbf{w})|. \tag{A.3}$$

A.2. Analytic functions

Theorem A.1. *Let* S_n, $n=0,1,\dots$ *be the operators denoted in* (2.12)*. A function* f *is analytic on* U_ρ *if and only if*

$$\limsup_{n\to\infty}\|S_n(f)\|_{L^2(I^q)}^{\frac{1}{n}}\le\rho. \tag{A.4}$$

Proof of Theorem A.1. Suppose f is analytic on U_ρ, then f is analytic on the closure of $U_{1/(\rho+\eta)}$. By (A.3) and (A.2),

$$\|S_n(f)\|_{L^2(I^q)}=\left(\sum_{|\mathbf{k}|_1=n}\left|\hat{f}(\mathbf{k})\right|^2\right)^{1/2}\le\sqrt{\binom{n+q-1}{q-1}}(\rho+\eta)^{|\mathbf{k}|_1}\max_{\mathbf{z}\in U_{1/(\rho+\eta)}}|f(\mathbf{z})|\le C(\eta)(\rho+2\eta)^n.$$

Thus,

$$\limsup_{n\to\infty} \|S_n(f)\|_{L^2(I^q)}^{1/n} \le \rho.$$

Now suppose $\|S_n(f)\|_{L^2(I^q)} \le \rho^n$ for all $n \in \mathbb{N}$, then

$$\left|\hat{f}(\mathbf{k})\right| \le \left(\sum_{|\mathbf{j}|_1=|\mathbf{k}|_1} \left|\hat{f}(\mathbf{j})\right|^2\right)^{1/2} = \left\|S_{|\mathbf{k}|_1}(f)\right\|_{L^2(I^q)} \le \rho^{|\mathbf{k}|_1}, \quad \mathbf{k}\in\mathbb{N}.$$

For any $\mathbf{z} \in U_{1/\rho}$, let $\rho' := \left(\max_{1\le j\le q} \left|z_j + \sqrt{z_j^2-1}\right|\right)^{-1} > \rho$, then (A.1) implies that

$$|p_{\mathbf{k}}(\mathbf{z})| \le \max_{\mathbf{w}\in\partial\Gamma_\rho} |g(\mathbf{w})| \prod_{j=1}^{q} \left|z_j + \sqrt{z_j^2-1}\right|^{k_j} \le \prod_{j=1}^{q} (1/\rho')^{k_j}.$$

Together with (A.3),

$$\sum_{\mathbf{k}\in\mathbb{N}^q} \left|\hat{f}(\mathbf{k})p_{\mathbf{k}}(\mathbf{z})\right| \le \max_{\mathbf{w}\in\partial U_\rho} |g(\mathbf{w})| \sum_{n=0}^{\infty}\sum_{|\mathbf{k}|_1=n} \frac{\rho^n}{\rho'^n} = \max_{\mathbf{w}\in\partial\Gamma_\rho} |g(\mathbf{w})| \sum_{n=0}^{\infty} \binom{n+q-1}{q-1}\left(\frac{\rho}{\rho'}\right)^n < \infty.$$

Hence, f is analytic at $\mathbf{z}$, which implies f is analytic on U_ρ. □

A.3. Entire functions

Theorem A.2. *Let $Q \in \mathbb{N}$, $\mathbf{v} = (v_1, \dots, v_Q)$, $\mathbf{r} = (r_1,\dots,r_Q) \in \mathbb{R}_+^Q$, $I_{\mathbf{r}} = \prod_{j=1}^{Q}[-r_j, r_j] \subset \mathbb{R}^Q$ and $\{p_{\mathbf{k},\mathbf{r}}\}_{\mathbf{k}\in\mathbb{N}^Q}$ the multivariable Chebyshev polynomials orthonormal on $I_{\mathbf{r}}$. Let $F: \mathbb{C}^Q \to \mathbb{C}$ be an entire function with*

$$\sup_{\mathbf{z}\in\mathbb{C}^Q} |F(\mathbf{z})| \le \exp\left\{\sum_{j=1}^{Q} v_j|z_j|\right\}, \tag{A.5}$$

then

$$\left\|\sum_{|\mathbf{k}|_1=N} \hat{F}_{\mathbf{r}}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}\right\|_{L_\infty(I_{\mathbf{r}})} \le 2\left(\frac{2\pi}{Q}\right)^{Q/2} N^{Q/2}\frac{(\mathbf{v}\cdot\mathbf{r})^N}{N!}. \tag{A.6}$$

Conversely, if F is a function on $\mathbb{C}^Q$ satisfying (A.6) for each $\mathbf{r}\in\mathbb{R}_+^Q$, then we can prove it is an entire function.

Theorem A.3. *Let $Q \in \mathbb{N}$, $\mathbf{v} = (v_1,\dots,v_Q) \in \mathbb{R}_+^Q$. If a function $F: \mathbb{C}^Q \to \mathbb{C}$ satisfies (A.6) for any $\mathbf{r}\in\mathbb{R}^Q$, then for any $\mathbf{z}\in\mathbb{C}^Q$,*

$$\sup_{\mathbf{z}\in\mathbb{C}^Q}\left(\frac{|F(\mathbf{z})|}{\exp\left\{2\left(\sum_{j=1}^{Q} v_j|z_j|\right)(1+\eta)\right\}}\right) < \infty, \quad \forall\eta > 0. \tag{A.7}$$

Proof of Theorem A.2. First, we consider $I_{\mathbf{r}} = I^q$ as the unit cube. In this case, we write $p_{\mathbf{k},\mathbf{r}}$ as $p_{\mathbf{k}}$ for $\mathbf{k}\in\mathbb{N}^Q$.
By (A.2), we have

$$\left|\hat{F}(\mathbf{k})\right| \le \frac{1}{\mathbf{k}^{\mathbf{k}}}\max_{\mathbf{z}\in U_\rho}|F(\mathbf{z})| \le \frac{e^{|\mathbf{k}|_1}\mathbf{v}^{\mathbf{k}}}{\mathbf{k}^{\mathbf{k}}}.$$

By Stirling's approximation,

$$\left|\hat{F}(\mathbf{k})\right| \le 2\frac{(\sqrt{2\pi})^Q(k_1\dots k_Q)^{1/2}(\mathbf{v}+\boldsymbol{\eta})^{\mathbf{k}}}{\mathbf{k}!}$$

for some A'_η depending only on η and Q.

Observing that

$$(y_1+\cdots+y_Q)^N=\sum_{\|\mathbf{k}\|_1=N}\frac{N!}{k_1!k_2!\dots k_Q!}\prod_{j=1}^{Q}y_j^{k_j}, \tag{A.8}$$

we have

$$\|S_N(F)\|_{L_\infty(I^Q)}=\sup_{\mathbf{x}\in I^q}\left|\sum_{|\mathbf{k}|_1=N}\hat{F}(\mathbf{k})P_{\mathbf{k}}(\mathbf{x})\right|\le\sum_{|\mathbf{k}|_1=N}\left|\hat{F}(\mathbf{k})\right|\le 2\left(\frac{2\pi}{Q}\right)^{Q/2}\frac{N^{Q/2}}{N!}|\mathbf{v}|_1^N.$$

Now we make a change of variables. For $\mathbf{r}\in\mathbb{R}_+^Q$, let $G(\mathbf{z})=F(\mathbf{z}\mathbf{r})$, then G is an entire function with

$$\sup_{\mathbf{z}\in\mathbb{C}^Q}|G(\mathbf{z})|\le\exp\left\{\sum_{j=1}^{Q}v_jr_j|z_j|\right\}.$$

Hence,

$$\left\|\sum_{|\mathbf{k}|_1=N}\hat{G}(\mathbf{k})p_{\mathbf{k}}\right\|_{L_\infty(I^q)}\le 2\left(\frac{2\pi}{Q}\right)^{Q/2}N^{Q/2}\frac{(\mathbf{v}\cdot\mathbf{r})^N}{N!},$$

where

$$\hat{G}(\mathbf{k})=\int_{I^q}G(\mathbf{x})p_{\mathbf{k}}(\mathbf{x})d\mathbf{x}=\int_{I^q}F(\mathbf{r}\circ\mathbf{x})p_{\mathbf{k}}(\mathbf{x})v_Q(\mathbf{x})d\mathbf{x}=\left(\prod_{j=0}^{Q}r_j\right)\int_{I_{\mathbf{r}}}F(\mathbf{y})p_{\mathbf{k}}\left(\frac{y_1}{r_1},\dots,\frac{y_Q}{r_Q}\right)v_{Q,\mathbf{r}}(\mathbf{y})\,d\mathbf{y}.$$

It is known that $\{p_{\mathbf{k},\mathbf{r}}\}_{\mathbf{k}\in\mathbb{N}^Q}$ is denoted by

$$p_{\mathbf{k},\mathbf{r}}(\mathbf{y})=p_{\mathbf{k}}\left(\frac{y_1}{r_1},\dots,\frac{y_Q}{r_Q}\right),\quad y\in I_{\mathbf{r}},$$

hence $\hat{F}_{\mathbf{r}}(\mathbf{k})=\hat{G}(\mathbf{k})$. Therefore,

$$\|S_N(F)\|_{L_\infty(I_{\mathbf{r}})}=\left\|\sum_{|\mathbf{k}|_1=N}\hat{F}_{\mathbf{r}}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}\right\|_{L_\infty(I_{\mathbf{r}})}\le\left\|\sum_{|\mathbf{k}|_1=N}\hat{G}(\mathbf{k})p_{\mathbf{k}}\right\|_{L_\infty(I^q)}\le 2\left(\frac{2\pi}{Q}\right)^{Q/2}N^{Q/2}\frac{(\mathbf{v}\cdot\mathbf{r})^N}{N!}.$$

This proves Theorem A.2. □

Proof of Theorem A.3. Suppose

$$\|S_N(F)\|_{L_\infty(I_{\mathbf{r}})}\le 2\left(\frac{2\pi}{Q}\right)^{Q/2}N^{Q/2}\frac{(\mathbf{v}\cdot\mathbf{r})^N}{N!}$$

holds true for every $\mathbf{r}\in\mathbb{R}^Q$.

For any $\mathbf{z}\in\mathbb{C}^Q$, take $\mathbf{r}=(|z_1|,\dots,|z_Q|)$. Since

$$\sup_{|z|=1,z\in\mathbb{C}}\frac{\left|z+\sqrt{z^2-1}\right|}{|z|}\le 2,$$

by (A.1),

$$|p_{\mathbf{k},\mathbf{r}}(\mathbf{z})|\le 2^N\|p_{\mathbf{k},\mathbf{r}}\|_{L_\infty(I_{\mathbf{r}})},\quad \mathbf{k}\in\mathbb{N}^Q.$$

With $C=2\left(\frac{2\pi}{Q}\right)^{Q/2}$,

$$\sum_{|\mathbf{k}|_1=N}\left|\hat{F}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}(\mathbf{z})\right|\le CN^{Q/2}\frac{(2\mathbf{v}\cdot\mathbf{r})^N}{N!}.$$

We will use Stirling's approximation to eliminate the $N^{Q/2}$ term. By Stirling's approximation, for $N\ge Q/2$,

$$\sum_{|\mathbf{k}|_1=N}\left|\hat{F}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}(\mathbf{z})\right| \le \frac{C}{\sqrt{2\pi N}}N^{Q/2}\left(\frac{e}{N}\right)^N(2\mathbf{v}\cdot\mathbf{r})^N \le \frac{C}{\sqrt{2\pi N}}\left(\frac{e}{N-Q/2}\right)^{N-Q/2}e^{Q/2}(2\mathbf{v}\cdot\mathbf{r})^N$$
$$\le \frac{C}{\sqrt{2\pi N}}\frac{2\sqrt{2\pi(N-Q/2)}}{(N-Q/2)!}e^{Q/2}(2\mathbf{v}\cdot\mathbf{r})^N$$
$$\le 2\left(\frac{2\pi}{Q}\right)^{Q/2}(e\mathbf{v}\cdot\mathbf{r})^{Q/2}\frac{(2\mathbf{v}\cdot\mathbf{r})^{N-Q/2}}{(N-Q/2)!}.$$

Therefore for $\mathbf{r}\in\mathbb{R}^Q_{\ge 1}$, $\sum_{|\mathbf{k}|_1<Q/2}\left|\hat{F}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}(\mathbf{z})\right|$ is bounded by $\mathcal{P}_1(2\mathbf{v}\cdot\mathbf{r})$ with $\mathcal{P}_1$ a polynomial of degree $Q/2-1$ and

$$\sum_{|\mathbf{k}|_1\ge Q/2}\left|\hat{F}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}(\mathbf{z})\right| \le \sum_{N=Q/2}^{\infty}2C(e\mathbf{v}\cdot\mathbf{r})^{Q/2}\frac{(2\mathbf{v}\cdot\mathbf{r})^{N-Q/2}}{(N-Q/2)!} = \sum_{N=0}^{\infty}2C(e\mathbf{v}\cdot\mathbf{r})^{Q/2}\frac{(2\mathbf{v}\cdot\mathbf{r})^{N}}{N!}$$
$$= 2C(e\mathbf{v}\cdot\mathbf{r})^{Q/2}\exp(2\mathbf{v}\cdot\mathbf{r}) \le 4\left(\frac{2\pi}{Q}\right)^{Q/2}(e\mathbf{v}\cdot\mathbf{r})^{Q/2}\exp(2\mathbf{v}\cdot\mathbf{r}).$$

Now we can bound $F(\mathbf{z})$ by

$$|F(\mathbf{z})| \le \sum_{|\mathbf{k}|_1<Q/2}\left|\hat{F}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}(\mathbf{z})\right| + \sum_{|\mathbf{k}|_1\ge Q/2}\left|\hat{F}(\mathbf{k})p_{\mathbf{k},\mathbf{r}}(\mathbf{z})\right| \le \mathcal{P}_1(2\mathbf{v}\cdot\mathbf{r}) + \mathcal{P}_2(2\mathbf{v}\cdot\mathbf{r})\exp(2\mathbf{v}\cdot\mathbf{r})$$

with $\mathcal{P}_2(x) = 4\left(\frac{2\pi}{Q}\right)^{Q/2}(ex)^{Q/2}$.

Since $\mathcal{P}_1$ and $\mathcal{P}_2$ are polynomials of degree at most Q, we conclude for any $\eta>0$, there exists some constant A_η depending on η and Q such that

$$|F(\mathbf{z})| \le A_\eta \exp(2(\mathbf{v}+\boldsymbol{\eta})\cdot\mathbf{r}(1+\eta)) = A_\eta \exp\left(\left(2\sum_{j=1}^{Q}(v_j+\eta)|z_j|\right)(1+\eta)\right). \quad \square$$

References

[1] A. Fisher, C. Rudin, F. Dominici, All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously, J. Mach. Learn. Res. 20 (177) (2019) 1–81.
[2] D.E. Knuth, The Art of Computer Programming, vol. 1, Addison Wesley, 1975.
[3] G.G. Lorentz, M. von Golitschek, Y. Makovoz, Constructive Approximation: Advanced Problems, vol. 304, Springer, Berlin, 1996.
[4] H.N. Mhaskar, Kernel-based analysis of massive data, Front. Appl. Math. Stat. 6 (2020) 30.
[5] L. Semenova, C. Rudin, R. Parr, A study in Rashomon curves and volumes: a new perspective on generalization and model simplicity in machine learning, arXiv preprint, arXiv:1908.01755, 2019.
[6] A. Zygmund, Trigonometric Series, vol. 1, Cambridge University Press, 2002.