ELSEVIER
Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®
Journal of Complexity 20 (2004) 561–590
Journal of COMPLEXITY
http://www.elsevier.com/locate/jco

On the tractability of multivariate integration and approximation by neural networks

H.N. Mhaskar[1]

Department of Mathematics, California State University, Los Angeles, CA 90032, USA

Received 4 June 2003; accepted 12 November 2003

Dedicated to Hans–Peter Blatt and Charles Micchelli in celebration of their 61st birthday

Abstract

Let $q \geqslant 1$ be an integer, Q be a Borel subset of the Euclidean space $\mathbb{R}^q$, μ be a probability measure on Q, and $\mathscr{F}$ be a class of real valued, μ-integrable functions on Q. The complexity problem of approximating $\int f \, d\mu$ using quasi-Monte Carlo methods is to estimate

$$\mathscr{E}_n(\mathscr{F}, \mu) := \inf_{x_1, \dots, x_n \in Q} \sup_{f \in \mathscr{F}} \left| \int f \, d\mu - \frac{1}{n} \sum_{k=1}^{n} f(x_k) \right|.$$

The problem is said to be tractable if there exist constants c, α, β independent of q (but possibly dependent on μ and $\mathscr{F}$) such that $\mathscr{E}_n(\mathscr{F}, \mu) \leqslant c q^{\alpha} n^{-\beta}$. We explore different regions (including manifolds), function classes, and measures for which this problem is tractable. Our results include tractability theorems for integration with respect to non-tensor product measures, and over unbounded and/or non-tensor product subsets, including the unit spheres of $\mathbb{R}^q$ with respect to various norms. We discuss applications to approximation capabilities of neural and radial basis function networks.

MSC: 65D30; 41A25; 41A55

Keywords: Tractability; Neural networks; Radial basis function networks; Dimension independent bounds

E-mail address: hmhaska@calstatela.edu.

[1] The research was supported, in part, by Grant DMS-0204704 from the National Science Foundation and Grant DAAD19-01-1-0001 from the U.S. Army Research Office.

0885-064X/$ - see front matter © 2003 Published by Elsevier Inc.
doi:10.1016/j.jco.2003.11.004

1. Introduction

In many applications, one needs to approximate the multi-dimensional integral $\int_Q f(\mathbf{x})\,d\mu(\mathbf{x})$ where Q is a Borel subset of a Euclidean space $\mathbb{R}^q$ (where $q\geqslant 1$ is an integer), and μ is a probability measure supported on Q. Such problems arise, for example, in mathematical finance [15,16], statistical learning theory [20], and approximation by neural and radial basis function networks [2,8–12]. The quasi-Monte Carlo technique for this approximation is to choose appropriate points $\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*$ so that the average of the values $f(\mathbf{x}_k^*)$ approximates this integral. An important question in complexity theory is to estimate

$$\mathscr{E}_n(\mathscr{F},\mu) := \inf_{\mathbf{x}_1,\ldots,\mathbf{x}_n\in Q}\ \sup_{f\in\mathscr{F}}\left|\int_Q f(\mathbf{x})\,d\mu(\mathbf{x}) - \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{x}_k)\right|, \tag{1.1}$$

for a suitable class $\mathscr{F}$ of functions. The problem is said to be *tractable* if

$$\mathscr{E}_n(\mathscr{F},\mu)\leqslant\frac{cq^\alpha}{n^\beta} \tag{1.2}$$

for some constants $c, \alpha, \beta > 0$, independent of q. However, the constants may depend upon μ and $\mathscr{F}$.

In addition to $\mathscr{E}_n(\mathscr{F},\mu)$, it is often customary to study the *normalized error*, defined by

$$\hat{\mathscr{E}}_n(\mathscr{F},\mu) = \frac{\mathscr{E}_n(\mathscr{F},\mu)}{\sup_{f\in\mathscr{F}}|\int f\,d\mu|}. \tag{1.3}$$

The denominator in the above fraction may be thought of as an ‘‘initial cost’’ or the ‘‘cost of doing nothing’’, and the normalized error measures the improvement of the quasi-Monte Carlo method over this initial cost. If $\mathscr{F}$ contains the function $\mathbb{1}$, which is identically equal to 1 on Q, then clearly, $\sup_{f\in\mathscr{F}}|\int f\,d\mu|\geqslant 1$. If, in addition, each function f in $\mathscr{F}$ satisfies $|f(x)|\leqslant 1$ $(x\in Q)$, then

$$\hat{\mathscr{E}}_n(\mathscr{F},\mu) = \mathscr{E}_n(\mathscr{F},\mu). \tag{1.4}$$

In many results on this subject, the tractability problem is studied for functions that can be represented in the form $\mathbf{x}\to\sigma(\Phi(\mathbf{x},\cdot))$, where Φ is a fixed kernel function (e.g., the reproducing kernel in a reproducing kernel Hilbert space), and σ varies over a suitable class of functionals. Novak and Woźniakowski have given two interesting surveys of this topic in [13,14].

Next, we note an interesting connection between the tractability problem for multivariate integration and approximation theory. Suppose that $\mathscr{F}$ is the unit ball of some normed linear function space X, on which point evaluations as well as the functional μ^*, given by $f\mapsto\int f\,d\mu$, are continuous linear functionals. If we denote the point evaluation functional at a point $\mathbf{x}$ by $\delta_\mathbf{x}$, then it is clear that $\mathscr{E}_n(\mathscr{F},\mu)$ gives an estimate on the degree of approximation of μ^* in the dual norm of the norm on X from the convex hull of $\{\delta_\mathbf{x}\}$. An important example of this line of thought, that includes both neural and radial basis function networks, is formulated in the following theorem.

Theorem 1.1. *Let Q, Q_1 be Borel subsets of some Euclidean spaces, $\Phi : Q \times Q_1 \to \mathbb{R}$ be a fixed, bounded Borel measurable kernel function, and $\mathscr{M}$ be a class of signed measures on Q having total variation equal to* 1. *We define $\mathscr{F}_\Sigma$ to be the class of all functions on Q of the form*

$$\mathbf{x} \mapsto \int_{Q_1} \Phi(\mathbf{x}, t)\, d\sigma(t), \tag{1.5}$$

where σ ranges over all signed measures on Q_1 having total variation 1, *and $\mathscr{F}_{\mathscr{M}}$ to be the class of all functions on Q_1 of the form $t \mapsto \int_Q \Phi(\mathbf{x}, t) d\mu(\mathbf{x})$, $\mu \in \mathscr{M}$. The following are equivalent*:

(a) *We have*

$$\mathscr{E}_n(\mathscr{F}_\Sigma, \mu) \leqslant \delta_n, \quad \mu \in \mathscr{M}. \tag{1.6}$$

(b) *We have*

$$\inf_{\mathbf{x}_1, \cdots, \mathbf{x}_n \in Q} \sup_{t \in Q_1} \left| g(t) - \frac{1}{n} \sum_{j=1}^{n} \Phi(\mathbf{x}_j, t) \right| \leqslant \delta_n, \quad g \in \mathscr{F}_{\mathscr{M}}. \tag{1.7}$$

Proof. Let (1.6) hold, $\mu \in \mathscr{M}$, and $\varepsilon > 0$ be arbitrary. Since $\Phi(\cdot, t) \in \mathscr{F}_\Sigma$ for every $t \in Q_1$, (1.6) implies that there exist $\mathbf{x}_1, \ldots, \mathbf{x}_n \in Q$ (independent of $t \in Q_1$) such that

$$\left| \int \Phi(\mathbf{x}, t)\, d\mu(\mathbf{x}) - \frac{1}{n} \sum_{j=1}^{n} \Phi(\mathbf{x}_j, t) \right| \leqslant \delta_n + \varepsilon, \quad t \in Q_1. \tag{1.8}$$

In view of the definition of the class $\mathscr{F}_{\mathscr{M}}$, this estimate is equivalent to estimate (1.7). Conversely, let (1.7) hold, $\mu \in \mathscr{M}$, and $\varepsilon > 0$ be arbitrary. Then there exist points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in Q$ such that (1.8) holds. Let $f(\mathbf{x}) = \int \Phi(\mathbf{x}, t) d\sigma(t)$ for some signed measure σ on Q_1 with total variation equal to 1. Since Φ is a bounded function, we may use Fubini's theorem to conclude that

$$\int \int \Phi(\mathbf{x}, t)\, d\mu(\mathbf{x})\, d\sigma(t) = \int \int \Phi(\mathbf{x}, t)\, d\sigma(t)\, d\mu(\mathbf{x}) = \int f(\mathbf{x})\, d\mu(\mathbf{x}).$$

Therefore, (1.8) leads to

$$\left| \int f(\mathbf{x})\, d\mu(\mathbf{x}) - \frac{1}{n} \sum_{j=1}^{n} f(\mathbf{x}_j) \right| \leqslant \delta_n + \varepsilon, \quad f \in \mathscr{F}_\Sigma,$$

i.e., $\mathscr{E}_n(\mathscr{F}_\Sigma, \mu) \leqslant \delta_n$. This proves (1.6). □

We observe that one needs an estimate of the form (1.6) uniformly for a large class of measures to make the approximation estimate (1.7) interesting.

Our first aim in this paper is to explore a general framework that enables us to analyse different regions, manifolds, function classes, and classes of measures for which an estimate of the form (1.2) can be obtained. Many of the known results on the tractability problem deal with "tensor product" function classes and measures. Our results include tractability theorems for integration with respect to non-tensor product measures, and over unbounded and/or non-tensor product subsets, including the unit spheres of $\mathbb{R}^q$ with respect to various norms.

Our second aim in this paper is to obtain bounds of the form cq^α/n^β on the degree of approximation by neural and radial basis function networks, where n is the number of "neurons" in the network (cf. Section 4 for the definition), and c, α, β are independent of n and q. Typically, the known estimates in this theory are of the form $c(q)/n^\beta$, where β is independent of q, but often without the requirement that $c(q)$ be polynomially dependent on q.

To give a preview of one of the novelties of our results in this paper, we recall, for example, that a radial basis function (RBF) network with activation function $\phi : [0, \infty) \to \mathbb{R}$ and n neurons (and norm $||\cdot||$) is a function of the form $\mathbf{x} \mapsto \sum_{j=1}^n a_j \phi(||\mathbf{x} - \mathbf{y}_j||)$, where $\mathbf{y}_j \in \mathbb{R}^q$ and $a_j \in \mathbb{R}$, $1 \leqslant j \leqslant n$. Most results on the degree of approximation by radial basis function networks assume the norm $||\cdot||$ to be the usual Euclidean norm. One novelty of our results is that we are able to supply some bounds in the case of any *absolute norm* (cf. Section 2.2 for the definition) in the argument of the activation function.

In the next section, we develop some basic concepts, which will be needed in formulating our results. The main theorems concerning integration are stated in Section 3. Section 4 describes some applications to the theory of approximation by neural and radial basis function networks. The proofs are given in Section 5.

2. Preparatory concepts

2.1. Measures

In this section, we introduce certain classes of measures which will be needed in the statement of our theorems.

We denote by λ_q the q-dimensional Lebesgue measure. For $1 \leqslant p \leqslant \infty$ and $\mathbf{x} \in \mathbb{R}^q$, we define

$$||\mathbf{x}||_p := ||\mathbf{x}||_{q,p} := \begin{cases} \left\{ \sum_{k=1}^{q} |x_k|^p \right\}^{1/p} & \text{if } 1 \leqslant p < \infty, \\ \max_{1 \leqslant k \leqslant q} |x_k| & \text{if } p = \infty. \end{cases} \tag{2.1}$$

Definition 2.1. Let μ be a probability measure on $\mathbb{R}^q$, $L, M, \beta, \gamma \geqslant 0$.

(a) The measure μ is said to satisfy a *decay condition* (with parameters (L, β)) if for all $\delta \in (0, 1]$,

$$\mu(\mathbb{R}^q \backslash [-L\delta^{-\beta}, L\delta^{-\beta}]^q) \leqslant \delta. \tag{2.2}$$

(b) The measure μ is said to satisfy a *continuity condition* (with parameters (M, γ)) if for all Borel sets $S \subseteq \mathbb{R}^q$,

$$\mu(S) \leqslant (M\lambda_q(S))^{\gamma}. \tag{2.3}$$

(c) The measure μ is said to be *regular* (with parameters (L, β, M, γ)) if it satisfies both (2.2) and (2.3).

(d) A signed measure σ will be called regular if the total variation measure $|\sigma|$ is regular in the sense of part (c). Similar terminology will apply for σ to satisfy a decay condition or a continuity condition.

Next, we give some examples of regular, non-tensor product measures.

Example 1. Let K be a compact subset of $\mathbb{R}^q$, $1 < p \leqslant \infty$, $f : K \to [0, \infty)$ be Lebesgue measurable, with a finite L^p norm N (with respect to the Lebesgue measure) on K, and $\int_K f \, d\lambda_q = 1$. The measure defined by $\mu_f(S) = \int_{S \cap K} f \, d\lambda_q$ is a regular measure with parameters $(\max_{\mathbf{x} \in K} ||\mathbf{x}||_{\infty}, 0, N^{p'}, 1/p')$, where $1/p + 1/p' = 1$.

Example 2. We give two examples of non-tensor product measures supported on the whole space. Let

$$\mu_{\text{exp}}(\alpha; S) := \lambda_{\text{exp},\alpha} \int_S \exp(-||\mathbf{x}||_2^{\alpha}) \, d\mathbf{x}, \tag{2.4}$$

where $\alpha > 0$, and

$$\lambda_{\text{exp},\alpha} := \frac{\alpha \Gamma(q/2)}{2\pi^{q/2} \Gamma(q/\alpha)} \tag{2.5}$$

is chosen to make $\mu_{\text{exp}}(\alpha; \mathbb{R}^q) = 1$ (cf. (5.33) below).

Another set of examples is given by

$$\mu_{\text{pow}}(\alpha; S) := \lambda_{\text{pow},\alpha} \int_S \frac{d\mathbf{x}}{1 + ||\mathbf{x}||_2^{\alpha}}, \tag{2.6}$$

where $\alpha > q$, and

$$\lambda_{\text{pow},\alpha} := \frac{\alpha \Gamma(q/2) \sin(\pi q/\alpha)}{2\pi^{(q+2)/2}} \tag{2.7}$$

is chosen to make $\mu_{\text{pow}}(\alpha; \mathbb{R}^q) = 1$.

Proposition 2.1. (a) *Let $\alpha > 0$. The measure $\mu_{\text{exp}}(\alpha)$ satisfies the continuity condition with $\gamma = 1$ and $M = \lambda_{\text{exp},\alpha}$. It satisfies the decay condition with any $\beta > 0$ and corresponding L given by*

$$L_{\text{exp}}^{\alpha} := \frac{2|q-\alpha|}{\alpha} + \left(\frac{1 + |q-\alpha|\beta}{\alpha\beta e}\right)^{1+|q-\alpha|\beta} \left(\frac{2}{\Gamma(q/\alpha)}\right)^{\alpha\beta} + 1. \tag{2.8}$$

(b) *Let* $\alpha > q$. *The measure* $\mu_{\rm pow}(\alpha)$ *satisfies the continuity condition with* $\gamma = 1$ *and* $M = \lambda_{{\rm pow},\alpha}$. *It satisfies the decay condition with* $\beta_{\rm pow} = 1/(\alpha - q)$ *and*

$$L_{\rm pow} := \left(\frac{\alpha \sin(\pi q/\alpha)}{\pi(\alpha - q)}\right)^{1/(\alpha-q)}. \tag{2.9}$$

By restricting and renormalizing these measures to different Borel sets, one can easily generate examples of regular non-tensor product measures supported on Borel sets other than the whole space, including sets that are both non-tensor product and unbounded.

2.2. Geometrical concepts

Let $||\cdot||$ be any absolute norm on $\mathbb{R}^q$, i.e., we assume that

$$||(x_1, \dots, x_q)|| = ||(|x_1|, \dots, |x_q|)||$$

for all $\mathbf{x} \in \mathbb{R}^q$. It is known (cf. [6, Theorem 5.5.10]) that $||\cdot||$ is *monotone*, i.e., $|x_j| \leqslant |y_j|$, $1 \leqslant j \leqslant q$, implies $||\mathbf{x}|| \leqslant ||\mathbf{y}||$. Let $\mathbf{e}_j$ be the unit vector whose jth component is 1 and other components are 0, $\kappa_1^{-1} := \min_{1 \leqslant j \leqslant q} ||\mathbf{e}_j||$ and $\kappa_2 := ||(1, \dots, 1)||$. Then the monotonicity of the norm leads to

$$\kappa_1^{-1} ||\mathbf{x}||_\infty \leqslant ||\mathbf{x}|| \leqslant \kappa_2 ||\mathbf{x}||_\infty, \quad \mathbf{x} \in \mathbb{R}^q. \tag{2.10}$$

In the sequel, we will adopt the following notation. If $\oplus$ is a binary operation on $\mathbb{R}$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^q$, then $\mathbf{x} \oplus \mathbf{y}$ will be the vector in $\mathbb{R}^q$ whose jth component is $x_j \oplus y_j$. If $c \in \mathbb{R}$, then $c \oplus \mathbf{x} := (c, \dots, c) \oplus \mathbf{x}$, and $\mathbf{x} \oplus c = \mathbf{x} \oplus (c, \dots, c)$. Conventions regarding the placement of the operator $\oplus$ will be continued as usual; for example, $\max(\mathbf{x}, \mathbf{y})$ is the vector whose jth component is $\max(x_j, y_j)$. Similar conventions are followed for binary relations. In particular, for $\mathbf{x} \in \mathbb{R}^q$, and $\mathbf{r} \in [0, \infty]^q$, we define the vector $\mathbf{z} = \frac{\mathbf{x}}{\mathbf{r}}$ by

$$z_j = \begin{cases} \infty & \text{if } r_j = 0 \text{ and } x_j \neq 0, \\ 0 & \text{if } r_j = x_j = 0, \\ 0 & \text{if } r_j = \infty, \\ \dfrac{x_j}{r_j} & \text{otherwise.} \end{cases}$$

If a component of $\mathbf{z}$ is infinity, we set $||\mathbf{z}|| := \infty$. For $\mathbf{y} \in \mathbb{R}^q$, $\mathbf{r} \in [0, \infty]^q$, we define the ellipsoid

$$B(||\cdot||, \mathbf{y}, \mathbf{r}) = \left\{\mathbf{x} \in \mathbb{R}^q : \left|\left|\frac{\mathbf{x} - \mathbf{y}}{\mathbf{r}}\right|\right| \leqslant 1\right\}. \tag{2.11}$$

We note that the values 0 and ∞ are both valid for the components of $\mathbf{r}$ in the above definition. If all components of $\mathbf{r}$ are equal to r, then the ellipsoid is the ball denoted by $B(||\cdot||, \mathbf{y}, r) := B(||\cdot||, \mathbf{y}, (r, \dots, r))$. We denote $B(||\cdot||, \mathbf{0}, 1)$ by $B_{||\cdot||}$, its volume by $\tau_{q,||\cdot||}$, its boundary by $\mathbb{S}^{q-1}_{||\cdot||}$, and the area of this boundary by $\omega_{q-1,||\cdot||}$. It is easy to see that

$$\lambda_q(B(||\cdot||, \mathbf{y}, \mathbf{r})) = \tau_{q,||\cdot||}\prod_{k=1}^{q} r_k, \tag{2.12}$$

where $0\cdot\infty := 0$.

Next, we introduce some notations concerning strips. Let $\mathbf{x}\cdot\mathbf{y}$ denote the inner product of $\mathbf{x}$ and $\mathbf{y}$. For $\mathbf{y}\in\mathbb{S}^{q-1}_{||\cdot||_2}$, and $a, b\in\mathbb{R}$, $a\leqslant b$, we write

$$S(\mathbf{y}, a, b) := \{\mathbf{x}\in\mathbb{R}^q : \mathbf{x}\cdot\mathbf{y}\in[a,b]\}. \tag{2.13}$$

Further, $S(\mathbf{y}, a, \infty) := \bigcup_{b\in(a,\infty)} S(\mathbf{y}, a, b)$, $S(\mathbf{y}, -\infty, b) := \bigcup_{a\in(-\infty,b)} S(\mathbf{y}, a, b)$. If $a>b$, we define $S(\mathbf{y}, a, b)$ to be the empty set.

Finally, we discuss some notation related to the sphere. A cap of radius r in $\mathbb{S}^{q-1}_{||\cdot||}$ centered at $\mathbf{y}$ is defined by

$$\mathbb{S}^{q-1}_{||\cdot||,r}(\mathbf{y}) := \{\mathbf{x}\in\mathbb{S}^{q-1}_{||\cdot||} : ||\mathbf{x}-\mathbf{y}||\leqslant r\}. \tag{2.14}$$

2.3. Function classes

For a subset $S\subseteq\mathbb{R}^q$, the characteristic function of S is defined by

$$\chi(S; x) := \begin{cases} 1 & \text{if } x\in S, \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

The constant function taking the value 1 everywhere on $\mathbb{R}^q$ will be denoted by $\mathbb{1}$.

An estimate on $\mathscr{E}_n(\mathscr{F}, \mu)$ where $\mathscr{F}$ consists of characteristic functions of certain sets is usually called a *discrepancy estimate*. In Section 5, we will obtain the discrepancy estimates for the following classes of characteristic functions.

We start with the set of characteristic functions of ellipsoids:

$$\mathscr{B}(||\cdot||, R, R_1) := \{\chi(B(||\cdot||, \mathbf{y}, \mathbf{r})) : \mathbf{y}\in\overline{(-R,R)^q},\ \mathbf{r}\in[0,R_1]^q\}\cup\{\mathbb{1}\}, \quad R, R_1\geqslant 0, \tag{2.16}$$

where R or R_1 may also be infinity. In the case of the norm $||\cdot||_\infty$, the ellipsoids are just cells in $\mathbb{R}^q$, and we write

$$\mathscr{R}(R, R_1) := \mathscr{B}(||\cdot||_\infty, R, R_1). \tag{2.17}$$

(We recall that an open cell in $\mathbb{R}^q$ is a set of the form $\prod_{k=1}^{q} I_k$, where each I_k is an open interval in $\mathbb{R}$. By a cell, we will mean the closure of an open cell.)

Similarly, we define

$$\mathscr{K}_{||\cdot||} := \{\chi(\mathbb{S}^{q-1}_{||\cdot||,r}(\mathbf{y})) : \mathbf{y}\in\mathbb{S}^{q-1}_{||\cdot||},\ r\geqslant 0\} \tag{2.18}$$

and

$$\mathscr{S}(R) := \{\chi(S(\mathbf{y}, a, b)\cap B(||\cdot||_2, \mathbf{0}, R)) : \mathbf{y}\in\mathbb{S}^{q-1}_{||\cdot||_2}, a, b\in\mathbb{R}\}\cup\{\mathbb{1}\}. \tag{2.19}$$

We note that the class $\mathscr{K}_{||\cdot||}$ already contains the function $\mathbb{1}$.

Clearly, any estimate on $\mathscr{E}_n(\mathscr{F},\mu)$ is valid also if $\mathscr{F}$ is replaced by its signed convex hull, i.e., the set of functions of the form $\sum a_j f_j$, where the sum is a finite sum, $f_j \in \mathscr{F}$ for each j, and a_j's are real numbers with $\sum |a_j| \leqslant 1$. We may write $\sum a_j f_j$ in the form $\int \Phi(\cdot, j)\, d\sigma(j)$, where for each j involved in the sum, $\Phi(\cdot, j) := f_j$, and σ is the signed measure that associates the mass a_j with the integer j. With this motivation in mind, we now proceed to define the notion of a *generalized convex hull* of $\mathscr{F}$, denoted by $\mathrm{conv}(\mathscr{F})$. In the case when $\mathscr{F}$ is the set of characteristic functions defined above, $\mathrm{conv}(\mathscr{F})$ contains functions of the form (2.21) or (2.22) (described below), as well as some other sets of functions recently considered in the literature on tractability problems.

Let $\mathscr{F}$ be a class of functions on a subset Q of $\mathbb{R}^q$, and Q_1 be a measure space. An $\mathscr{F}$ valued process on Q_1 is a jointly measurable mapping $\Phi : Q \times Q_1 \to \mathbb{R}$, such that for each $t \in Q_1$, $\Phi(\cdot, t) \in \mathscr{F}$. The *generalized convex hull* of $\mathscr{F}$ with respect to Q_1, denoted by $\mathrm{conv}(\mathscr{F}, Q_1)$, is defined to be the set of all functions of the form

$$\mathbf{x} \mapsto \int \Phi(\mathbf{x}, t)\, d\sigma(t), \tag{2.20}$$

where σ ranges over all signed measures on Q_1 having total variation not exceeding 1, and Φ ranges over all $\mathscr{F}$ valued processes on Q_1. The class $\mathrm{conv}(\mathscr{F})$ consists of functions that are in $\mathrm{conv}(\mathscr{F}, Q_1)$ for some measure space Q_1. (Here, we have tacitly assumed a "universal set" of all measure spaces of interest. In this paper, this universal set consists of all Borel measurable subsets of all finite-dimensional Euclidean spaces.)

Next, we discuss some examples of the notion of generalized convex hulls. We recall (cf. [18, Chapter 8, Sections 12–21]) that there is a one-to-one correspondence between signed measures having bounded variation on $\mathbb{R}$ and functions having bounded variation on $\mathbb{R}$. Thus, if $\phi : \mathbb{R} \to \mathbb{R}$ is a right (respectively, left) continuous function having bounded variation, and $\phi(x) \to 0$ as $x \to -\infty$ (respectively, $\phi(x) \to 0$ as $x \to \infty$), then there exists a unique signed measure μ_ϕ such that $\phi(x) = \mu_\phi((-\infty, x])$ (respectively, $\phi(x) = -\mu_\phi([x, \infty))$) for all $x \in \mathbb{R}$, and the total variation of this measure is the same as the total variation of ϕ. Similar representations hold for functions defined on subintervals of $\mathbb{R}$, satisfying different one-sided continuity conditions, and normalizations. Therefore, one usually thinks of ϕ itself as a signed measure, and writes $d\phi$ in place of $d\mu_\phi$, where μ_ϕ is the measure appropriate to the normalizations of ϕ. The corresponding total variation measure is usually denoted (in the context of integrations) by $|d\phi|$.

Example 3. Let ϕ be a left continuous function of bounded variation on $[0, \infty)$, such that $\lim_{x\to\infty} \phi(x) = 0$, and the total variation of ϕ is equal to 1. If σ is a signed measure on $\mathbb{R}^q \times [0, \infty)^q$, having total variation equal to 1, then a function of the form

$$\mathbf{x} \mapsto \int \phi\left(\left\|\frac{\mathbf{x}-\mathbf{y}}{\mathbf{r}}\right\|\right) d\sigma(\mathbf{y}, \mathbf{r}) = -\int\int \chi(B(\|\cdot\|, \mathbf{x}, \mathbf{r}u); \mathbf{y})\, d\phi(u)\, d\sigma(\mathbf{y}, \mathbf{r}) \tag{2.21}$$

is in $\mathrm{conv}(\mathscr{B}(\|\cdot\|, \infty, \infty))$.

Example 4. Let $\phi : \mathbb{R} \to [0, \infty)$ be a right continuous function of bounded variation with $\lim_{x\to-\infty} \phi(x) = 0$, σ be a signed measure on $\mathbb{R}^q$, having total variation equal to 1. Then a function of the form

$$\mathbf{x} \mapsto \int \phi(\mathbf{x}\cdot\mathbf{y})\, d\sigma(\mathbf{y}) = \begin{cases} \phi(0)\sigma(\mathbb{R}^q) & \text{if } \mathbf{x} = \mathbf{0}, \\ \int\int \chi(S(\mathbf{x}/\|\mathbf{x}\|, u/\|\mathbf{x}\|, \infty); \mathbf{y})\, d\phi(u)\, d\sigma(\mathbf{y}) & \text{if } \mathbf{x} \neq \mathbf{0} \end{cases} \tag{2.22}$$

is in $\text{conv}(\mathscr{S}(\infty))$. Functions of the forms described in this and the previous example are of interest in the theory of neural networks and radial basis function networks, respectively. We will examine these in further detail in Section 4.

Example 5. In this example, we discuss the set $\text{conv}(\mathscr{R}(R, R_1))$ in some detail. In [5], Hickernell, Sloan, and Wasilkowski have studied the tractability of quasi-Monte Carlo approximation of an integral of the form $\int_Q F(\mathbf{x})W(\mathbf{x})\, d\mathbf{x}$, where Q is a (bounded or unbounded) cell in $\mathbb{R}^q$, and $W(\mathbf{x}) = \prod_{k=1}^q W_k(x_k)$ for some weights $W : \mathbb{R} \to [0, \infty)$. They have pointed out that by simple substitutions, this problem is equivalent to the problem of approximating $\int_D f(\mathbf{x})\, d\mathbf{x}$, where $D = [-1/2, 1/2]^q$.

The problem is proved to be tractable for the class $\mathscr{F}_H$ of functions, defined as follows. We fix an *anchor* $\mathbf{c} \in D$. For a subset $U \subseteq \{1, \dots, q\}$, let $D_U \coloneqq [-1/2, 1/2]^{|U|}$. For $\mathbf{x} \in \mathbb{R}^q$, let $\mathbf{x}_U$ denote the vector of length $|U|$ whose components are the components x_j of $\mathbf{x}$ for which $j \in U$, and $(\mathbf{x}_U, \mathbf{c})$ be the q-dimensional vector whose kth component is x_k if $k \in U$, and c_k if $k \notin U$. For a sufficiently smooth function $f : D \to \mathbb{R}$ to allow the following differentiation, we write

$$f'_U(\mathbf{x}_U) = \frac{\partial^{|U|}}{\prod_{k\in U} \partial x_k} f(\mathbf{x}_U, \mathbf{c}). \tag{2.23}$$

If U is the empty set, the corresponding f'_U is defined to be the constant function $f(\mathbf{c})$. If U is the empty set, it is also convenient to define $\|f'_U\|_{1,D_U} \coloneqq |f(\mathbf{c})|$. The class $\mathscr{F}_H$ consists of all functions $f : D \to \mathbb{R}$ for which $f'_U(\mathbf{x}_U)$ exists for each $\mathbf{x} \in [-1/2, 1/2]^q$ and for each $U \subseteq \{1, \dots, q\}$, and

$$\|f\|_{1,D} \coloneqq \sum_{U\subseteq\{1,\dots,q\}} \int_{D_U} |f'(\mathbf{x}_U)|\, d\mathbf{x}_U \leqslant 1. \tag{2.24}$$

For $f \in \mathscr{F}_H$ and $\mathbf{x} \geqslant \mathbf{c}$, we have the integral representation (cf. [5])

$$f(\mathbf{x}) = \sum_{U\subseteq\{1,\dots,q\}} \int_{[0,1/2]^{|U|}} \chi([(\mathbf{y}_U, \mathbf{c}), (1/2, \dots, 1/2)]; \mathbf{x}) f'_U(\mathbf{y}_U)\, d\mathbf{y}_U. \tag{2.25}$$

Similar representations hold in each of the 2^q quadrants of D defined by $\mathbf{c}$. We choose the anchor $\mathbf{c} = \mathbf{0}$, and observe that each of the cells involved in (2.25) (and its analogue in the other quadrants) has its center in $[-1/2, 1/2]^q$ and $\|\cdot\|_\infty$-radius not exceeding 1/4. Thus, the class $\mathscr{F}_H$ is seen to be a subset of $\text{conv}(\mathscr{R}(1/2, 1/4))$. Hickernell, Sloan, and Wasilkowski have already made use of this observation in [5] to deduce an estimate on $\mathscr{E}_n(\mathscr{F}_H, \lambda_q)$ from that on the class of characteristic functions of cells. The additional observation here is regarding the radii and locations of centers of the cells involved.

The following proposition summarizes an observation which we will use extensively.

Proposition 2.2. *Let μ be a probability measure on a Borel measurable subset Q of $\mathbb{R}^q$, and $\mathscr{F}$ be a class of Borel measurable functions on Q, such that $|f(x)| \leqslant 1$ for all $f \in \mathscr{F}$ and $x \in Q$. We have for integer $n \geqslant 1$,*

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{F}), \mu) = \mathscr{E}_n(\mathscr{F}, \mu) = \inf_{\mathbf{x}_1, \ldots, \mathbf{x}_n \in Q} \sup_{f \in \mathscr{F}} \left| \int_Q f(\mathbf{x})\, d\mu(\mathbf{x}) - \frac{1}{n} \sum_{k=1}^{n} f(\mathbf{x}_k) \right|. \tag{2.26}$$

In particular, if $\varepsilon > 0$, one may choose points $\mathbf{x}_j$ depending only on ε, μ and $\mathscr{F}$, and independently of the measure spaces, processes, and measures needed to define functions in $\mathrm{conv}(\mathscr{F})$, *such that*

$$\sup_{f \in \mathrm{conv}(\mathscr{F})} \left| \int_Q f(\mathbf{x})\, d\mu(\mathbf{x}) - \frac{1}{n} \sum_{k=1}^{n} f(\mathbf{x}_k) \right| \leqslant \mathscr{E}_n(\mathscr{F}, \mu) + \varepsilon.$$

3. Tractability of integration

In this section, we will discuss a variety of theorems estimating $\mathscr{E}_n(\mathscr{F}, \mu)$ for different function classes and measures. In each case, $\mathscr{F}$ includes the function $\mathbb{1}$, and $|f| \leqslant 1$ for all $f \in \mathscr{F}$. Hence, the normalized error $\hat{\mathscr{E}}_n(\mathscr{F}, \mu) = \mathscr{E}_n(\mathscr{F}, \mu)$ in each case. In the sequel, we write

$$G := \frac{4}{3\log 3 - 2} \approx 3.0868, \tag{3.1}$$

and, for $\kappa, B > 0$,

$$\Delta_n(\kappa, B) := 2\sqrt{\frac{G}{n}\{B + (\kappa/2)\log(n/(GB))\}}. \tag{3.2}$$

In general, our estimates will have the form $\mathscr{E}_n(\mathscr{F}, \mu) \leqslant \Delta_n(\kappa, B)$, where the constants κ, B will be given explicitly, in terms of the various parameters defining the function classes and the decay/continuity conditions on the measures. It is perhaps possible to sharpen these results with a removal of the term $\log n$, using ideas from V-C theory of probability, as in [5]. However, this is expected to give an unspecified constant depending on $\mathscr{F}$ and μ. We have decided to choose explicitly defined constants, even if they might not be the best ones, and also have the slightly weaker result with the logarithmic term, in order to make it easier to determine whether our theorems imply tractability for particular measures and function classes, with absolute constants.

Our first theorem is an extension of estimate (21) in [5, Theorem 3] of Hickernell, Sloan, and Wasilkowski regarding the tractability of integration with respect to λ_q on $[-1/2, 1/2]^q$.

Theorem 3.1. *Let* $0 < R, R_1 < \infty$, $M, \gamma > 0$, *and* $L, \beta \geqslant 0$.

(a) *Let*

$$2qM(4R_1)^{q-1}\min(3R_1/4, R) \geqslant 1. \tag{3.3}$$

With B defined in (5.10), *we have for* $n \geqslant GB$, *and any measure* μ *satisfying a continuity condition with parameters* (M, γ),

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{R}(R, R_1)), \mu) \leqslant \varDelta_n(2q/\gamma, B). \tag{3.4}$$

(b) *Let*

$$Mq\tau_{q,||\cdot||}(2R_1)^{q-1}\min(1, 3R_1/4, \sqrt{R}(1 + \kappa_2 + \kappa_2 R_1)) \geqslant 1. \tag{3.5}$$

With B defined in (5.15), *we have for* $n \geqslant GB$, *and any measure* μ *satisfying a continuity condition with parameters* (M, γ),

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{B}(||\cdot||, R, R_1)), \mu) \leqslant \varDelta_n(3q\gamma^{-1}, B). \tag{3.6}$$

(c) *Let* $qM(2^{2+\beta}L)^q \geqslant 2$. *With the constant B* as in (5.12), *we have for* $n \geqslant GB$, *and any regular measure* μ *with parameters* (L, β, M, γ),

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{R}(\infty, \infty)), \mu) \leqslant \varDelta_n(2q(\beta q + 1/\gamma), B). \tag{3.7}$$

Part (b) of this theorem is clearly a generalization of part (a), except for different constants. We present part (a) separately to allow a comparison with the result in [5] (Example 6 below). We note that the support of the measure μ in part (c) may well be an unbounded and non-tensor product set. In the most general cases, the value of B determines the tractability, and is $\mathcal{O}(q^2 \log q)$, where the constant involved in $\mathcal{O}$ may depend upon μ, R, R_1, and the norm $||\cdot||$. In some special cases, however, the value of B is smaller. We illustrate this with a few examples.

Example 6. Theorem 3.1(a) may be applied to the case explained in Example 5. In this example only, let $D = [-1/2, 1/2]^q$, $M \in [1, \infty)$, $w : [-1/2, 1/2]^q \to [0, M]$, and $\int_D w(\mathbf{x})\, d\mathbf{x} = 1$. The measure μ defined on D by $d\mu = w(\mathbf{x})\, d\mathbf{x}$ satisfies a continuity condition with parameters $(M, 1)$. Since $\mathscr{F}_H \subset \mathrm{conv}(\mathscr{R}(1/2, 1/4))$, we take $R = 1/2$, and $R_1 = 1/4$. Since $M \geqslant 1$, condition (3.3) is satisfied if $q \geqslant 3$. Part (a) of the above theorem therefore implies that for the class $\mathscr{F}_H$ (with the anchor fixed at $\mathbf{0}$), we have $B = (4q + 1)\log 2 + 2q\log(qM)$, and

$$\mathscr{E}_n(\mathscr{F}_H, \mu) \leqslant 2\left\{\frac{G}{n}(B + q\log(n/(GB)))\right\}^{1/2}. \tag{3.8}$$

We note that w does not need to be a tensor product function. In the case when $w \equiv 1$, we recover the corresponding result in [5] as far as the order of magnitude of the dependence on q and n is concerned, apart from the values of the different constants involved.

Example 7. The purpose of this example is to illustrate Theorem 3.1(b) with a non-tensor product region of integration and non-tensor product measure. We take $||\cdot|| = ||\cdot||_2$, and omit the reference to this norm from the notations. To avoid conflict of notation, we write $U = B_{||\cdot||}$. Let (in this example only), Φ denote the class of functions $\phi : [0,\infty)\to[0,1]$ which are decreasing on $[0,\infty)$, with $\phi(0) = 1$, and $\phi(x) = 0$ if $x\geqslant 1/2$, $\mathscr{M}$ be the set of all measures supported on U having total variation 1, and $\mathscr{F}$ be the class of functions of the form

$$\begin{aligned} f(\mathbf{x}) &= \int \phi(||\mathbf{x}-\mathbf{y}||)\, d\sigma(\mathbf{y}) \\ &= -\int\int \chi(B(||\cdot||,\mathbf{x},u);\mathbf{y})\, d\phi(u)\, d\sigma(\mathbf{y}), \quad \mathbf{x}\in U, \quad \phi\in\Phi, \quad \sigma\in\mathscr{M}. \end{aligned}$$

We observe that $\mathscr{F}\subset\mathrm{conv}(\mathscr{B}(||\cdot||,1,1/2))$. Let $g:[0,1]\to[0,\infty)$ be an integrable function with $\int_0^1 g(u)\,du = 1$, $0\leqslant g(u)\leqslant N$ for $u\in[0,1]$, and μ be the measure defined on U by $d\mu(\mathbf{x}) = \tau_q^{-1} g(||\mathbf{x}||^q)\, d\lambda_q(\mathbf{x})$. Then μ is a probability measure, satisfying a continuity condition with $M = N/\tau_q$, $\gamma = 1$. Condition (3.5) is satisfied if $q\geqslant 3$. The value of B in (5.15) is given by $(4q+1)\log 2 + 3q\log N + 2q\log(2+3\sqrt{q}) + 3q\log q$.

We observe that if a measure satisfies a decay condition with parameters (L,β), then it also satisfies a decay condition with parameters (L_1,β_1) for any $L_1\geqslant L$ and $\beta_1\geqslant\beta$. Similarly, if it satisfies a continuity condition with parameters (M,γ) then it also satisfies a continuity condition with parameters (M_1,γ) for all $M_1\geqslant M$. Therefore, condition (3.3) may be omitted by replacing M in (5.10) by $\max(M,\{2q(4R_1)^{q-1}\min(3R_1/4,R)\}^{-1})$. Similar remarks hold also for conditions (3.5), the lower bound condition in Theorem 3.1(c), and other similar conditions in the other theorems in this paper. However, we feel that the formulations given here allows better clarity in the different formulas as well as better flexibility in applying the results.

Our next theorem deals with tractability on the spheres.

Theorem 3.2. *Let* $M_1\geqslant 1$, $\gamma_1>0$ *and the constant* B *be defined by* (5.17). *For any integer* $n\geqslant GB$ *and any measure* μ, *supported on* $\mathbb{S}^{q-1}_{||\cdot||}$ *and satisfying a spherical continuity condition,*

$$\sup_{\mathbf{y}\in\mathbb{S}^{q-1}_{||\cdot||}} \mu(\mathbb{S}^{q-1}_{||\cdot||,r}(\mathbf{y})\backslash\mathbb{S}^{q-1}_{||\cdot||,\rho}(\mathbf{y}))\leqslant (M_1(r-\rho))^{\gamma_1}, \quad r\geqslant\rho\geqslant 0, \tag{3.9}$$

we have

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{K}_{||\cdot||}),\mu)\leqslant \Delta_n(q/\gamma_1,B). \tag{3.10}$$

The constant B in (5.17) is $\mathcal{O}(q)$, although the constants involved may depend upon both μ and $||\cdot||$. We elaborate upon an example, which we find especially interesting.

Example 8. Let K be any compact, convex subset of $\mathbb{R}^q$, $\mathbf{0}$ be in the interior of K, and K be symmetric in the sense that $\mathbf{x} \in K$ if and only if $(|x_1|, \ldots, |x_k|) \in K$. The *Minkowski functional* for K is defined by

$$||\mathbf{x}||_K := \inf\{t > 0 : t^{-1}\mathbf{x} \in K\}. \tag{3.11}$$

It is well known ([6, Theorem 5.5.8 and its proof]) that $||\cdot||_K$ is an absolute norm, and $K = B_{||\cdot||_K}$. Conversely, for any absolute norm $||\cdot||$, $||\cdot|| = ||\cdot||_{B_{||\cdot||}}$. In particular, if the unit vectors $\mathbf{e_j}$ are on the boundary of K, then

$$||\mathbf{x}||_\infty \leqslant ||\mathbf{x}||_K \leqslant ||\mathbf{x}||_1 \leqslant q||\mathbf{x}||_\infty. \tag{3.12}$$

Thus, Theorem 3.2 applies to integration over sets lying on the boundary of sets K satisfying the properties mentioned above. The constant B in (3.10) in such cases is $q \log q + q \log(7M_1) + 3 \log 2$.

Finally, we state a theorem related to classes defined in terms of strips.

Theorem 3.3. *Let* $R, \gamma, M > 0$, $L, \beta \geqslant 0$.

(a) *Let* $2M\tau_{q-1,||\cdot||_{q-1,2}}R^q \geqslant 1$. *With* B *as in* (5.20), *we have for integer* $n \geqslant GB$, *and any measure* μ *satisfying a continuity condition with parameters* (M, γ),

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{S}(R)), \mu) \leqslant \Delta_n((q+1)/\gamma, B). \tag{3.13}$$

(b) *Let*

$$2^{\beta q+1}\tau_{q-1,||\cdot||_{q-1,2}}ML^q \geqslant 1. \tag{3.14}$$

With B *as in* (5.22), *we have for integer* $n \geqslant GB$, *and any regular measure* μ *with parameters* (L, β, M, γ),

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{S}(\infty)), \mu) \leqslant \Delta_n((q+1)(q\beta + 1/\gamma), B). \tag{3.15}$$

In the above theorem, as usual, $B = \mathcal{O}(q^2 \log q)$, although the constants may depend on μ and R. An important class of functions for which Theorem 3.3 implies tractability in this sense is the class of all functions of the form

$$\mathbf{x} \mapsto \int \exp(-\mathbf{x} \cdot \mathbf{y})\, d\sigma(\mathbf{y}), \quad \mathbf{x} \geqslant 0,$$

for a signed measure σ on $[0, \infty)^q$ with total variation equal to 1. Every function in this class has an analytic extension to the right half-plane with respect to each of the components of $\mathbf{x}$. If σ is a probability measure, the function is completely monotone in each of its variables.

4. Applications

In this section, we discuss certain applications of the theorems in Section 3 to the theory of neural networks and radial basis function networks.

4.1. Neural networks

Let $\phi : \mathbb{R} \to \mathbb{R}$. A neural network with activation function ϕ, and having n neurons, is a function of the form $\mathbf{x} \mapsto \sum_{j=1}^{n} c_j \phi(\mathbf{x} \cdot \mathbf{w}_j + b_j)$, where the *output layer weights* $c_j \in \mathbb{R}$, the *synaptic weights* $\mathbf{w}_j \in \mathbb{R}^q$, and the *thresholds* $b_j \in \mathbb{R}$. The theory of approximation by neural networks is quite well developed (cf. [11] for a survey in the context of approximation of classical Sobolev classes). In [1], Barron has studied functions on $\mathbb{R}^{q-1}$ which can be expressed in the form

$$F(\mathbf{x}) = \int_{\mathbb{S}^{q-1}_{\|\cdot\|_2} \times [-1,1]} \chi([0, \infty); \mathbf{x} \cdot \mathbf{y} + r) G(\mathbf{y}, r) \, d\sigma_1(\mathbf{y}, r),$$

where σ_1 is the product measure of the normalized area measure on $\mathbb{S}^{q-1}_{\|\cdot\|_2}$ with the one-dimensional Lebesgue measure, and

$$\int_{\mathbb{S}^{q-1}_{\|\cdot\|_2} \times [-1,1]} |G(\mathbf{y}, r)| \, d\sigma_1(\mathbf{y}, r) = 1.$$

For such functions, he proved that for any integer $n \geqslant 1$, there exist $a_k, b_k \in [-1, 1]$, $\mathbf{y}_k \in \mathbb{S}^{q-1}_{\|\cdot\|_2}$, $1 \leqslant k \leqslant n$, such that

$$\max_{\mathbf{x} \in B_{\|\cdot\|_2}} \left| F(\mathbf{x}) - \sum_{k=1}^{n} a_k \chi([0, \infty); \mathbf{x} \cdot \mathbf{y}_k + b_k) \right| \leqslant c \sqrt{\log n / n},$$

where c is a constant depending only on q. Similar results have been proved by many authors, [2,7–10,12]. In particular, Kurkova [7] has given bounds in a Hilbert space setting that depend polynomially on q. We observe that we may define a function on $\mathbb{R}^q$ by the formula

$$f(\mathbf{x}) = \int_{\mathbb{S}^{q-1}_{\|\cdot\|_2} \times [-1,1]} \chi([0, \infty); \mathbf{x} \cdot \mathbf{y}) g(\mathbf{y}) \, d\sigma_2(\mathbf{y}),$$

where g and σ_2 are just G and σ_1 expressed in a different notation. Then $f((x_1, \ldots, x_{q-1}, 1)) = F((x_1, \ldots, x_{q-1}))$.

Motivated by this example, we define the following class of functions. Let ϕ be a function having bounded variation on $\mathbb{R}$. The class $\mathscr{F}_{\mathscr{N}}(\phi, L, \beta, M, \gamma)$ consists of all functions of the form

$$\mathbf{x} \mapsto \int \phi(\mathbf{x} \cdot \mathbf{y}) \, d\mu(\mathbf{y}),$$

where μ is a regular, signed measure with parameters (L, β, M, γ).

Theorem 4.1. *Suppose ϕ is a function having bounded variation on $\mathbb{R}$, with the normalizations that ϕ is right continuous,* $\lim_{x\to-\infty}\phi(x)=0$, *and the total variation of ϕ is* 1. *Let $f\in\mathscr{F}_{\mathscr{N}}(\phi,L,\beta,M,\gamma)$, where condition* (3.14) *is satisfied, and B be the constant defined in* (5.22). *Then for integer $n\geqslant GB$, there exist points $\mathbf{y}_j\in\mathbb{R}^q$, $j=1,\dots,n$,* (*depending on f*) *such that*

$$\sup_{\mathbf{x}\in\mathbb{R}^q}\left|f(\mathbf{x})-\frac{1}{n}\sum_{j=1}^{n}\phi(\mathbf{x}\cdot\mathbf{y}_j)\right|\leqslant\Delta_n((q+1)(q\beta+1/\gamma),B). \tag{4.1}$$

We note again that $B=\mathcal{O}(q^2\log q)$. In addition, all the output layer weights in our network are equal to $1/n$. The proof of Theorem 4.1 is a simple consequence of Proposition 5.5(b). We are also able to place bounds on the synaptic weights $\mathbf{y}_j$, provided the measure μ in the definition of the target function f, as well as the function ϕ, are compactly supported, and the approximation is desired on a compact set. This is done using Proposition 5.5(a), but no new ideas are needed. As far as we are aware, Theorem 4.1 is the first of its kind, where the degree of uniform approximation by neural networks on the whole Euclidean space is estimated.

4.2. Radial basis function networks

Let $\phi:[0,\infty)\to\mathbb{R}$. A radial basis function (RBF) network with activation function ϕ and n neurons (and norm $||\cdot||$) is a function of the form $\mathbf{x}\mapsto\sum_{j=1}^{n}a_j\phi(||\mathbf{x}-\mathbf{y}_j||)$, where the *centers* $\mathbf{y}_j\in\mathbb{R}^q$ and the *weights* $a_j\in\mathbb{R}$, $1\leqslant j\leqslant n$. Approximation by RBF networks has also been very popular in different applications, ranging from pattern recognition to the production of animated cartoons.

In this subsection, the function $\phi:[0,\infty)\to\mathbb{R}$ is a function having bounded variation on $[0,\infty)$, with the normalizations that ϕ is left continuous, $\lim_{x\to\infty}\phi(x)=0$, and the total variation of ϕ is 1. We assume further that ϕ satisfies the decay condition

$$\int_{L\delta^{-\beta}}^{\infty}|d\phi(x)|\leqslant\delta,\quad 0<\delta\leqslant 1. \tag{4.2}$$

We now consider the class $\mathscr{F}_{\mathscr{R}}(\phi,||\cdot||,L,\beta,M,\gamma)$ consisting of functions of the form

$$\mathbf{x}\mapsto\int\phi(||\mathbf{x}-\mathbf{y}||)\,d\mu(\mathbf{y}),$$

where μ is a regular, signed measure with parameters (L,β,M,γ). We note that there is no loss of generality in assuming that the same L and β are used here as in (4.2). Requiring the two values to be different will only result in a more elaborate book-keeping, but not in new ideas. This class is analogous to the "native space" for the function ϕ.

Theorem 4.2. *Let $L,\beta,M,\gamma>0$. We define $B_n=B$ as in* (5.15) *with $R_1=L(\log n/n)^{-\beta/2}$, $R=\kappa_1(1+\kappa_2)R_1$, and $2^{1/\gamma}M$ in place of M. Let n be sufficiently* *large, so that* $n \geqslant GB_n$, $\log n/n \leqslant 1/4$, *and condition* (3.5) *is satisfied with these parameters. Then for* $f \in \mathscr{F}_{\mathscr{R}}(\phi, ||\cdot||, L, \beta, M, \gamma)$, *there exist* $\mathbf{y}_j = \mathbf{y}_j(f) \in [-R_1, R_1]^q$ *such that*

$$\sup_{\mathbf{x} \in \mathbb{R}^q} \left| f(\mathbf{x}) - \frac{1}{n} \sum_{j=1}^{n} \phi(||\mathbf{x} - \mathbf{y}_j||) \right| \leqslant \Delta_n(3q\gamma^{-1}, B_n) + 7\sqrt{\frac{\log n}{n}}. \tag{4.3}$$

We note that $B_n = \mathcal{O}(q^2 \log(qn))$.

As far as we are aware, this is the first result of its kind where uniform approximation bounds are obtained for RBF networks using a norm other than the Euclidean norm on $\mathbb{R}^q$. It appears to be the first result of its kind proving a tractability result for uniform approximation by RBF networks on the entire Euclidean space. It is amusing to note that the weights in our networks are again all equal to $1/n$. Our proof can be modified to yield analogous results where the centers $\mathbf{y}_j$ are restricted to a compact cell in $\mathbb{R}^q$, and the approximation is also desired on a compact cell. We do not feel that this adds any new ideas.

5. Proofs

For clarity of presentation, we postpone the proof of Proposition 2.1 until the end of this section. The results in Section 4 are simple applications of those in Section 3. Our strategy for proving the theorems in Section 3 is as follows. In light of Proposition 2.2, it is enough to estimate $\mathscr{E}_n(\mathscr{F}, \mu)$ when $\mathscr{F}$ is the set of characteristic functions of the sets involved in each theorem. We will use the geometrical properties of the sets and the notion of one-sided entropy (“entropy with brackets” in the terminology in the book [19] of van der Vaart and Wellner) to obtain a finite set Y of characteristic functions such that $\mathscr{E}_n(\mathscr{F}, \mu)$ can be estimated using $\mathscr{E}_n(Y, \mu)$. This process is codified in Theorem 5.1 below. The problem of estimating $\mathscr{E}_n(\mathscr{F}, \mu)$ thus reduces to estimating the one-sided entropy of $\mathscr{F}$. The details of this estimation depend heavily on the geometrical properties of the sets, and we had to present them in the form of different propositions, in spite of a common theme behind all these estimations.

Before proving other theorems, we prove Proposition 2.2.

Proof of Proposition 2.2. It is clear that $\mathscr{F} \subseteq \mathrm{conv}(\mathscr{F})$, so that

$$\mathscr{E}_n(\mathrm{conv}(\mathscr{F}), \mu) \geqslant \mathscr{E}_n(\mathscr{F}, \mu).$$

In the proof of the reverse inequality, we note that $\mathscr{E}_n(\mathscr{F}, \mu) < \infty$. Let $\varepsilon > 0$ be arbitrary and $\mathbf{x}_j$ be chosen so that

$$\sup_{g \in \mathscr{F}} \left| \int g(\mathbf{x})\, d\mu(\mathbf{x}) - \frac{1}{n} \sum_{k=1}^{n} g(\mathbf{x}_k) \right| \leqslant \mathscr{E}_n(\mathscr{F}, \mu) + \varepsilon. \tag{5.1}$$

Let $f \in \mathrm{conv}(\mathscr{F})$. There exists an $\mathscr{F}$-valued process Φ and a signed measure σ of total variation 1 such that

$$f(\mathbf{x}) = \int \Phi(\mathbf{x}, t)\, d\sigma(t), \quad \mathbf{x} \in Q.$$

Using Fubini's theorem, we see that

$$\int f(\mathbf{x})\, d\mu(\mathbf{x}) - \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{x}_k) = \int \left\{ \int \Phi(\mathbf{x}, t)\, d\mu(\mathbf{x}) - \frac{1}{n}\sum_{k=1}^{n} \Phi(\mathbf{x}_k, t) \right\} d\sigma(t).$$

Since $\Phi(\cdot, t) \in \mathscr{F}$ for each t, we conclude from (5.1) that

$$\sup_{f \in \mathrm{conv}(\mathscr{F})} \left| \int f(\mathbf{x})\, d\mu(\mathbf{x}) - \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{x}_k) \right| \leqslant \mathscr{E}_n(\mathscr{F}, \mu) + \varepsilon.$$

Since ε is arbitrary, this completes the proof. □

We now begin with the proofs of the theorems in Sections 3 and 4. Towards this end, we define the notion of one-sided entropy, and prove a general estimate for quantities of the form $\mathscr{E}_n(\mathscr{F}, \mu)$ in terms of this one-sided entropy. Let Q be a measure space, μ be a probability measure defined on Q, $\mathscr{F}$ be a class of μ-integrable functions on Q, and $\delta > 0$. A finite set Y of μ-integrable functions on Q is said to be a one-sided (μ, δ)-cover of $\mathscr{F}$ if for every $f \in \mathscr{F}$, there exist $g, h \in Y$ with $g \leqslant f \leqslant h$ *everywhere* on Q, and $\int_Q (h - g)\, d\mu \leqslant \delta$. We observe that Y need not be a subset of $\mathscr{F}$. If $N(\mathscr{F}, \mu, \delta)$ is the number of elements in a minimal one-sided (μ, δ)-cover of $\mathscr{F}$, then we define the *one-sided entropy* $H(\mathscr{F}, \mu, \delta)$ to be the quantity $\log N(\mathscr{F}, \mu, \delta)$, where we find it convenient to take the natural logarithm.

The starting point of our investigations is the following observation. It is probably known in the statistical literature, but we find it easier to prove it than finding a reference.

Theorem 5.1. *Let (Q, μ) be a probability space, and $\mathscr{F}$ be a set of real valued, μ-integrable functions on Q, such that $|f(x)| \leqslant 1$ for all $f \in \mathscr{F}$ and $x \in Q$, and the one-sided entropy $H(\mathscr{F}, \mu, \cdot)$ satisfies*

$$H(\mathscr{F}, \mu, \delta) \leqslant \log A - \kappa \log \delta, \quad 0 < \delta \leqslant 1, \tag{5.2}$$

for some positive constants A and κ depending on $\mathscr{F}$, Q, and μ. Let $B := \log(2A)$. Then for any integer $n \geqslant GB$, there exist a set $T \subseteq Q$, consisting of n points, such that

$$\left| \int f\, d\mu - \frac{1}{n}\sum_{t \in T} f(t) \right| \leqslant \Delta_n(\kappa, B) = 2\sqrt{\frac{G}{n}\{B + (\kappa/2)\log(n/(GB))\}}, \tag{5.3}$$

where G is the constant defined in (3.1).

The proof of Theorem (5.1) mimics an argument in [4]. The main ingredient is to use the following sharper version of the Hoeffding's inequality (cf. [17, p. 191]). It is proved in [5], but not stated in this way.

Proposition 5.1. *Let* (Q,μ) *be a probability space,* $n\geqslant 1$ *be an integer, and* $\{X_k\}$, $k=1,\dots,n$ *be independent random variables on* Q*, each with range contained in a compact interval* $[a,b]$ *and expectation equal to* m*. Then for any* $\varepsilon\in(0,(b-a)/2]$,

$$\mathrm{Prob}\left(\left|n^{-1}\sum_{k=1}^{n} X_k - m\right|\geqslant\varepsilon\right)\leqslant 2\exp\left(-\frac{4n\varepsilon^2}{G(b-a)^2}\right). \tag{5.4}$$

Proof. Let $Z_j := (X_j-a)/(b-a)$, $1\leqslant j\leqslant n$. Then for $1\leqslant j\leqslant n$, we have $0\leqslant Z_j\leqslant 1$, the expected value of Z_j is $(m-a)/(b-a)$, and the variance of Z_j can be estimated by

$$\int Z_j^2\,d\mu-\left(\int Z_j\,d\mu\right)^2\leqslant\int Z_j\,d\mu-\left(\int Z_j\,d\mu\right)^2\leqslant 1/4.$$

Following [5], we now recall the Bennett inequality [17, p. 192]. According to this inequality, if Y_j are independent random variables, each with mean 0, range in $[-M,M]$, and variance σ_j, and $V\geqslant\sum_{j=1}^{n}\sigma_j^2$, then for $\eta>0$,

$$\mathrm{Prob}\left(\left|\sum_{j=1}^{n} Y_j\right|\geqslant\eta\right)\leqslant 2\exp\left(-\frac{V}{M^2}g(M\eta/V)\right), \tag{5.5}$$

where, in this proof only, $g(t) := (1+t)\log(1+t)-t$. We apply this estimate with $Y_j = Z_j-(m-a)/(b-a)$. Then we may choose $M=1$, $V=n/4$ and $\eta = n\varepsilon/(b-a)$. This leads to

$$\begin{aligned}\mathrm{Prob}&\left(\left|n^{-1}\sum_{k=1}^{n} X_k - m\right|\geqslant\varepsilon\right)\\ &=\mathrm{Prob}\left(\left|\sum_{k=1}^{n} Z_k - n(m-a)/(b-a)\right|\geqslant\eta\right)\\ &\leqslant 2\exp\left(-\frac{n}{4}g(4\eta/n)\right).\end{aligned} \tag{5.6}$$

Using elementary calculus, one verifies (cf. [5]) that $g(t)\geqslant(3\log 3-2)t^2/4 = t^2/G$ if $t\in[0,2]$. Hence, if $0\leqslant\eta\leqslant n/2$; i.e., $\varepsilon\leqslant(b-a)/2$, then

$$\mathrm{Prob}\left(\left|n^{-1}\sum_{k=1}^{n} X_k - m\right|\geqslant\varepsilon\right)\leqslant 2\exp\left(-\frac{4}{nG}\eta^2\right) = 2\exp\left(-\frac{4n\varepsilon^2}{G(b-a)^2}\right).$$

This completes the proof. □

Proof of Theorem 5.1. If $n\leqslant G\{B+(\kappa/2)\log(n/(GB))\}$, then (5.3) is trivial. Therefore, in the remainder of this proof, we will assume that $n>G\{B+(\kappa/2)\log(n/(GB))\}$, and write $\delta := \Delta_n(\kappa,B)/2$. Our assumption that $n\geqslant GB$ implies that $n>GB$ and $\delta\in(0,1)$. Let Y be a minimal one sided (μ,δ)-cover for $\mathscr{F}$. By replacing each $g \in Y$ by the function

$$g_1(x) := \begin{cases} g(x) & \text{if } |g(x)| \leqslant 1, \\ 1 & \text{if } g(x) \geqslant 1, \\ -1 & \text{if } g(x) \leqslant -1, \end{cases}$$

we may assume without loss of generality that the functions $g \in Y$ satisfy $|g| \leqslant 1$ as well. Now, let $f \in \mathscr{F}$. Then there exist $g_1, g_2 \in Y$ such that $g_1 \leqslant f \leqslant g_2$ and $\int (g_2 - g_1)\, d\mu \leqslant \delta$. Then for any measure v on Q, $\int g_1\, dv \leqslant \int f\, dv \leqslant \int g_2\, dv$, and

$$\int g_2\, d\mu - \int g_2\, dv - \delta \leqslant \int f\, d\mu - \int f\, dv \leqslant \int g_1\, d\mu - \int g_1\, dv + \delta.$$

Consequently,

$$\sup_{f \in \mathscr{F}} \left| \int f\, d\mu - \int f\, dv \right| \leqslant \max_{g \in Y} \left| \int g\, d\mu - \int g\, dv \right| + \delta. \tag{5.7}$$

Now, let $g \in Y$. Following [4], we take a random sample ξ_k from Q, distributed according to μ, and consider the random variable $X_k = g(\xi_k)$. Then the expected value of X_k is $\int g\, d\mu$ and $|X_k| \leqslant 1$. Since $\delta \in (0, 1)$, Proposition 5.1 implies that

$$\text{Prob}\left(\left| \frac{1}{n} \sum_{k=1}^{n} g(\xi_k) - \int g\, d\mu \right| \geqslant \delta \right) \leqslant 2 \exp(-n\delta^2/G).$$

Hence,

$$\begin{aligned} &\text{Prob}\left(\max_{g \in Y} \left| \frac{1}{n} \sum_{k=1}^{n} g(\xi_k) - \int g\, d\mu \right| \geqslant \delta \right) \\ &\quad \leqslant 2|Y| \exp(-n\delta^2/G) \\ &\quad = \exp\left(\log 2 + H(\mathscr{F}, \mu, \delta) - n\delta^2/G\right) \\ &\quad \leqslant \exp\left(\log(2A) - \kappa \log \delta - n\delta^2/G\right) = \exp\left(B - \kappa \log \delta - n\delta^2/G\right) \\ &\quad = \exp\left(-\frac{\kappa}{2} \log\left(1 + \frac{\kappa}{2B} \log\left(\frac{n}{GB} \right) \right) \right) \\ &\quad < 1. \end{aligned} \tag{5.8}$$

Therefore, there exist points ξ_k such that

$$\max_{g \in Y} \left| \frac{1}{n} \sum_{k=1}^{n} g(\xi_k) - \int g\, d\mu \right| \leqslant \sqrt{\frac{G}{n} \left\{ B - (\kappa/2) \log \frac{GB}{n} \right\}}.$$

Along with (5.7) (with v being the measure that associates the mass $1/n$ with each ξ_k), this proves (5.3) (with $T = \{\xi_k\}$). $\square$

We now begin the program of estimating the one-sided entropies of the different sets of characteristic functions described in Section 2.3. The following simple estimate will be used often in this process. In the sequel, μ will denote a probability measure on $\mathbb{R}^q$.

Lemma 5.1. *If* $\eta > 0$, $0 < R < \infty$, *and* $\max_{1 \leqslant j \leqslant q} \max(|x_j + \eta|, |x_j|) \leqslant R$, *then*

$$\left| \prod_{j=1}^{q} (x_j + \eta) - \prod_{j=1}^{q} x_j \right| \leqslant q\eta R^{q-1}. \tag{5.9}$$

Proof. Estimate (5.9) follows immediately from the identity

$$\prod_{j=1}^{q} (x_j + \eta) - \prod_{j=1}^{q} x_j = \sum_{k=1}^{q} \left(\prod_{j=k}^{q} (x_j + \eta) \prod_{j=1}^{k-1} x_j - \prod_{j=k+1}^{q} (x_j + \eta) \prod_{j=1}^{k} x_j \right)$$
$$= \eta \sum_{k=1}^{q} \prod_{j=k+1}^{q} (x_j + \eta) \prod_{j=1}^{k-1} x_j. \quad \square$$

In order to prove Theorem 3.1, we first prove two propositions, Propositions 5.2 and 5.3.

Proposition 5.2. (a) *Let* $0 < R$, $R_1 < \infty$, *and* μ *be a measure satisfying a continuity condition with parameters* (M, γ). *Suppose that* (3.3) *is satisfied. With*

$$B = \log\left\{ 2(2qM)^{2q} \left(\frac{2R}{R_1} \right)^{q} (4R_1)^{2q^2} \right\}, \tag{5.10}$$

we have for $n \geqslant GB$,

$$\mathscr{E}_n(\mathscr{R}(R, R_1), \mu) \leqslant \Delta_n(2q/\gamma, B). \tag{5.11}$$

(b) *Let* μ *be a regular measure with parameters* (L, β, M, γ), *where* $qM(2^{2+\beta}L)^q \geqslant 2$. *With*

$$B = (2q^2(\beta + 2) + (3 + 2/\gamma)q + 2)\log 2 + 2q \log(qML^q), \tag{5.12}$$

we have, for $n \geqslant GB$,

$$\mathscr{E}_n(\mathscr{R}(\infty, \infty), \mu) \leqslant \Delta_n(2q(\beta q + 1/\gamma), B). \tag{5.13}$$

Proof. First, we prove part (a). Let $0 < \delta \leqslant 1$. In view of (3.3), there exists an integer $m \geqslant 3$ in the interval

$$[6qM(4R_1)^{q-1}R\delta^{-1/\gamma}, 8qM(4R_1)^{q-1}R\delta^{-1/\gamma}].$$

We divide the cube $[-R, R]^q$ into m^q congruent subcubes, and let (in this proof only) $\mathscr{C}$ denote the set of centers of these subcubes. Next, let $m_1 = R_1 m/R$. Condition (3.3) ensures that $m_1 \geqslant 4$. For $\mathbf{z} \in \mathscr{C}$ and multi-integer $\mathbf{k} \geqslant 1$, let $g_{\mathbf{z},\mathbf{k}}$ denote the characteristic function of the cell $B(\|\cdot\|_\infty, \mathbf{z}, \mathbf{k}R_1/m_1)$. If any component of $\mathbf{k}$ is not positive, we define $g_{\mathbf{z},\mathbf{k}} = 0$. The set consisting of $\mathbb{1}$ and the functions $g_{\mathbf{z},\mathbf{k}}$, $\mathbf{z} \in \mathscr{C}$, $0 \leqslant \mathbf{k} \leqslant m_1 + 2$ ($\mathbf{k} \in \mathbb{Z}^q$) will be denoted by $Y_\delta(R, R_1)$. Now, if $\mathbf{y} \in [-R, R]^q$ and $\mathbf{r} \in [0, R_1]^q$, then there exist $\mathbf{z} \in \mathscr{C}$ and multi-integer $\mathbf{k}$ with $0 \leqslant \mathbf{k} \leqslant m_1$, such that $\|\mathbf{y} - \mathbf{z}\|_\infty \leqslant R/m$, and $\mathbf{k}R_1/m_1 \leqslant \mathbf{r} \leqslant (\mathbf{k} + 1)R_1/m_1$. Denoting the characteristic function of $B(\|\cdot\|_\infty, \mathbf{y}, \mathbf{r})$ by

f, it is easy to verify that $g_{\mathbf{z},\mathbf{k}-1} \leqslant f \leqslant g_{\mathbf{z},\mathbf{k}+2}$. Further,

$$\begin{aligned}&\int (g_{\mathbf{z},\mathbf{k}+2} - g_{\mathbf{z},\mathbf{k}-1})\, d\lambda_q \\ &\leqslant \left(\frac{2R_1}{m_1}\right)^q \left[\prod_{j=1}^{q} (k_j+2) - \prod_{j=1}^{q} \max(0, k_j - 1)\right] \\ &\leqslant 3q\left(\frac{2R_1}{m_1}\right)^q (m_1+2)^{q-1} \leqslant 3q2^{q-1}(2R_1)^q/m_1 \\ &= 3q2^q(2R_1)^{q-1}R/m = 6qR(4R_1)^{q-1}/m \leqslant \delta^{1/\gamma}/M.\end{aligned}$$

Therefore, the continuity condition on μ implies that

$$\int (g_{\mathbf{z},\mathbf{k}+2} - g_{\mathbf{z},\mathbf{k}-1})\, d\mu \leqslant \delta.$$

Thus, the set $Y_\delta(R, R_1)$ is a one-sided (μ, δ)-cover of $\mathscr{R}(R, R_1)$. Therefore,

$$\begin{aligned}\exp(H(\mathscr{R}(R, R_1), \mu, \delta)) &\leqslant |Y_\delta(R, R_1)| \leqslant m^q(m_1+3)^q + 1 \\ &\leqslant m^q(m_1 + 3 + 1/m)^q \leqslant m^q(m_1 + 3 + 1/3)^q \\ &\leqslant 2^q m^{2q}(R_1/R)^q \\ &\leqslant 2^q(R_1/R)^q\{8qMR(4R_1)^{q-1}\delta^{-1/\gamma}\}^{2q} \\ &= (2qM)^{2q}\left(\frac{2R}{R_1}\right)^q (4R_1)^{2q^2}\delta^{-2q/\gamma}.\end{aligned} \tag{5.14}$$

In view of Theorem 5.1, this leads to (5.11).

To prove part (b), we let h be the characteristic function of $\mathbb{R}^q \backslash [-L(\delta/2)^{-\beta}, L(\delta/2)^{-\beta}]$, $R = R_1 = L(\delta/2)^{-\beta}$, and

$$Y = Y_{\delta/2}(R, R) \cup \{g + h : g \in Y_{\delta/2}(R, R)\}.$$

Now, for any $\mathbf{y} \in \mathbb{R}^q$ and $\mathbf{r} \geqslant 0$, $B(\|\cdot\|_\infty, \mathbf{y}, \mathbf{r}) \cap [-R, R]^q$ is either empty or equal to $B(\|\cdot\|_\infty, \mathbf{x}, \mathbf{r}_1)$ for some $\mathbf{x} \in [-R, R]^q$ and $\mathbf{r}_1 \in [0, R]^q$. Thus, any $f \in \mathscr{R}(\infty, \infty)$ can be expressed in the form $f = f_1 + f_2$, where $f_1 \in \mathscr{R}(R, R) \cup \{1 - \mathbb{I}\}$, and $0 \leqslant f_2 \leqslant h$. We may find $g_1, g_2 \in Y_{\delta/2}(R, R)$ such that $g_1 \leqslant f_1 \leqslant g_2$ and $\int (g_2 - g_1)\, d\mu \leqslant \delta/2$. Hence, $g_1 \leqslant f \leqslant g_2 + h$, and the decay condition for μ implies that

$$\int (g_2 + h - g_1)\, d\mu \leqslant \delta.$$

Thus, Y is a one-sided (μ, δ)-cover for $\mathscr{R}(\infty, \infty)$. The cardinality of Y is at most twice that of $Y_{\delta/2}(R, R)$. We substitute the values of $R = R_1$ in (5.14), and use $\delta/2$ in place of δ to deduce that

$$\begin{aligned}H(\mathscr{R}(\infty, \infty), \mu, \delta) &\leqslant \{(2(\beta+2)q^2 + (3 + 2/\gamma)q + 1\} \log 2 \\ &\quad + 2q\log(qML^q) - (2q(q\beta + 1/\gamma)) \log \delta.\end{aligned}$$

This estimate and Theorem 5.1 leads to (5.13). □

Proposition 5.3. *Let* μ *be a probability measure satisfying a continuity condition with parameters* (M,γ), $0<R, R_1<\infty$, *and* (3.5) *be satisfied. With*

$$B=\log\{2((4Mq\tau_{q,\|\cdot\|})^3R(1+\kappa_2+\kappa_2R_1)^2(2R_1)^{3q-2})^q\} \tag{5.15}$$

we have for integer $n\geqslant GB$,

$$\mathscr{E}_n(\mathscr{B}(\|\cdot\|,R,R_1),\mu)\leqslant\Delta_n(3q\gamma^{-1},B). \tag{5.16}$$

The next lemma supplies a detail required in the proof of this proposition.

Lemma 5.2. *Let* $\mathbf{r}\in[0,R_1]^q$, $0<\varepsilon\leqslant1$, $\|\mathbf{z}\|_\infty\leqslant\varepsilon^2$, *and* $\mathbf{r}'\geqslant\mathbf{r}+\varepsilon(1+\kappa_2+\kappa_2R_1)$. *Then* $B(\|\cdot\|,\mathbf{y},\mathbf{r})\subseteq B(\|\cdot\|,\mathbf{y}+\mathbf{z},\mathbf{r}')$.

Proof. Since $\|\cdot\|$ is monotone, $B(\|\cdot\|,\mathbf{y},\mathbf{r})\subseteq B(\|\cdot\|,\mathbf{y},\mathbf{r}+\varepsilon)$. Further,

$$\left\|\frac{\mathbf{x}-\mathbf{y}}{\mathbf{r}+\varepsilon}-\frac{\mathbf{x}-\mathbf{y}-\mathbf{z}}{\mathbf{r}+\varepsilon}\right\|\leqslant\frac{\|\mathbf{z}\|}{\varepsilon}\leqslant\kappa_2\varepsilon.$$

So, $\mathbf{x}\in B(\|\cdot\|,\mathbf{y},\mathbf{r}+\varepsilon)$ implies that $\mathbf{x}\in B(\|\cdot\|,\mathbf{y}+\mathbf{z},(\mathbf{r}+\varepsilon)(1+\kappa_2\varepsilon))$. Since

$$(\mathbf{r}+\varepsilon)(1+\kappa_2\varepsilon)\leqslant\mathbf{r}+\varepsilon+\kappa_2\varepsilon(R_1+\varepsilon)\leqslant\mathbf{r}+\varepsilon(1+\kappa_2+\kappa_2R_1)\leqslant\mathbf{r}',$$

the monotonicity of $\|\cdot\|$ implies that $\mathbf{x}\in B(\|\cdot\|,\mathbf{y}+\mathbf{z},\mathbf{r}')$. □

Proof of Proposition 5.3. In this proof, we will denote $\tau_{q,\|\cdot\|}$ by τ_q. We will estimate the one-sided entropy $H(\mathscr{B}(\|\cdot\|,R,R_1),\mu,\delta)$, and use Theorem 5.1. Let $\delta\in(0,1]$, and $m\geqslant\max(3,\sqrt{R})$ be an integer in the range

$$[3Mq\tau_q(2R_1)^{q-1}(1+\kappa_2+\kappa_2R_1)\sqrt{R}\delta^{-1/\gamma},$$
$$4Mq\tau_q(2R_1)^{q-1}(1+\kappa_2+\kappa_2R_1)\sqrt{R}\delta^{-1/\gamma}].$$

(Condition (3.5) ensures that such an integer exists for every $\delta\in(0,1]$.) We divide $[-R,R]^q$ into m^{2q} congruent subcubes, and let (in this proof only) $\mathscr{C}$ denote the set of centers of these subcubes. Let

$$m_1=\frac{R_1m}{(1+\kappa_2+\kappa_2R_1)\sqrt{R}}.$$

Again, Condition (3.5) implies that $m_1\geqslant4$. For $\mathbf{z}\in\mathscr{C}$ and multi-integer $\mathbf{k}$ with $1\leqslant\mathbf{k}\leqslant m_1+2$, let $g_{\mathbf{z},\mathbf{k}}$ denote the characteristic function of the ellipse $B(\|\cdot\|,\mathbf{z},\mathbf{k}R_1/m_1)$. If some component of a multi-integer $\mathbf{k}$ is not positive, we define $g_{\mathbf{z},\mathbf{k}}=0$. The set consisting of $\mathbb{1}$ and the functions $g_{\mathbf{z},\mathbf{k}}$, $0\leqslant\mathbf{k}\leqslant m_1+2$ ($\mathbf{k}\in\mathbb{Z}^q$) will be denoted by Y. Let $\mathbf{y}\in[-R,R]^q$, $\mathbf{r}\in[0,R_1]^q$, and f be the characteristic function of $B(\|\cdot\|,\mathbf{y},\mathbf{r})$. Then there exists $\mathbf{z}\in\mathscr{C}$ and multi-integer $\mathbf{k}$ with $0\leqslant\mathbf{k}\leqslant m_1$ such that $\|\mathbf{y}-\mathbf{z}\|_\infty\leqslant R/m^2$ and $\mathbf{k}R_1/m_1\leqslant\mathbf{r}\leqslant(\mathbf{k}+1)R_1/m_1$. Using Lemma 5.2, it is easy to verify that $g_{\mathbf{z},\mathbf{k}-1} \leqslant f \leqslant g_{\mathbf{z},\mathbf{k}+2}$. We observe that

$$\begin{aligned}\int (g_{\mathbf{z},\mathbf{k}+2} - g_{\mathbf{z},\mathbf{k}-1})\, d\lambda_q &\leqslant \tau_q \left(\frac{R_1}{m_1}\right)^q \left(\prod_{j=1}^{q} (k_j+2) - \prod_{j=1}^{q} \max(0, k_j - 1)\right)\\ &\leqslant 3q\tau_q \left(\frac{R_1}{m_1}\right)^q (m_1+2)^{q-1}\\ &\leqslant 3q\tau_q (2R_1)^{q-1}(1+\kappa_2+\kappa_2 R_1)\sqrt{R}/m \leqslant \delta^{1/\gamma}/M.\end{aligned}$$

Since $g_{\mathbf{z},\mathbf{k}+2} - g_{\mathbf{z},\mathbf{k}-1}$ is the characteristic function of a Borel measurable set, the continuity condition on μ implies that

$$\int (g_{\mathbf{z},\mathbf{k}+2} - g_{\mathbf{z},\mathbf{k}-1})\, d\mu \leqslant \delta.$$

Thus, the set Y is a one-sided (μ, δ)-cover for $\mathscr{B}(||\cdot||, R, R_1)$. The cardinality of Y is at most

$$m^{2q}(m_1+3)^q + 1 \leqslant m^{2q}(2m_1)^q = (2R_1(\sqrt{R}(1+\kappa_2+\kappa_2 R_1))^{-1})^q m^{3q}.$$

Recalling that $m \leqslant 4qM\tau_q(2R_1)^{q-1}(1+\kappa_2+\kappa_2 R_1)\sqrt{R}\delta^{-1/\gamma}$, the above estimate leads to

$$H(\mathscr{B}(||\cdot||, R, R_1), \mu, \delta) \leqslant \log((4Mq\tau_q)^3 R(1+\kappa_2+\kappa_2 R_1)^2 \times (2R_1)^{3q-2})^q - \frac{3q}{\gamma}\log\delta.$$

Along with Theorem 5.1, this leads to (5.16). □

Proof of Theorem 3.1. We recall (2.26). Parts (a) and (c) follow from Proposition 5.2, parts (a) and (b), respectively. Part (b) follows from Proposition 5.3. □

The proof of Theorem 3.2 requires the following proposition.

Proposition 5.4. *Let μ be a probability measure on $\mathbb{S}^{q-1}_{||\cdot||}$ satisfying the spherical continuity condition* (3.9), *where we assume further that $M_1 \geqslant 1$. Let*

$$B = \log\left\{\frac{8q}{\kappa_1\kappa_2}(7\kappa_1\kappa_2 M_1)^q\right\}. \quad (5.17)$$

Then for integer $n \geqslant GB$,

$$\mathscr{E}_n(\mathscr{K}_{||\cdot||}, \mu) \leqslant \Delta_n(q/\gamma_1, B). \quad (5.18)$$

Proof. It is easy to verify that for $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{q-1}_{||\cdot||_\infty}$,

$$||\mathbf{x}-\mathbf{y}||_\infty \leqslant 2\kappa_1\kappa_2 \left|\left|\frac{\mathbf{x}}{||\mathbf{x}||} - \frac{\mathbf{y}}{||\mathbf{y}||}\right|\right| \leqslant (2\kappa_1\kappa_2)^2 ||\mathbf{x}-\mathbf{y}||_\infty. \quad (5.19)$$

Let $\delta\in(0,1]$. Since $M_1\geqslant 1$, we may find an integer $m\geqslant 6\kappa_1\kappa_2$ in the interval

$$[6\kappa_1\kappa_2 M_1\delta^{-1/\gamma_1}, 7\kappa_1\kappa_2 M_1\delta^{-1/\gamma_1}].$$

We divide each of the $2q$ faces of $\mathbb{S}^{q-1}_{\|\cdot\|_\infty}$ into m^{q-1} congruent subcubes, and let $\mathscr{C}$ be the set of projections of the centers of these subcubes on $\mathbb{S}^{q-1}_{\|\cdot\|}$. For integer ℓ, let $r_\ell = 2\kappa_1\kappa_2\ell/m$. Let $g_{\mathbf{z},\ell}$ denote the characteristic function of the cap $\mathbb{S}^{q-1}_{\|\cdot\|,r_\ell}(\mathbf{z})$, $\mathbf{z}\in\mathscr{C}$, $1\leqslant\ell\leqslant m/(\kappa_1\kappa_2)+2$, ($\ell$ integer), and $g_{\mathbf{z},\ell}=0$, if $\ell\leqslant 0$. Let Y be the set of functions $g_{\mathbf{z},\ell}$, $\mathbf{z}\in\mathscr{C}$, $0\leqslant\ell\leqslant m/(\kappa_1\kappa_2)+2$. Now, let f be the characteristic function of $\mathbb{S}^{q}_{\|\cdot\|,r}(\mathbf{y})$. In view of (5.19), there exist $\mathbf{z}\in\mathscr{C}$ and integer k with $0\leqslant k\leqslant m/(\kappa_1\kappa_2)$ such that $\|\mathbf{y}-\mathbf{z}\|\leqslant 2\kappa_1\kappa_2/m$ and $r_k\leqslant r\leqslant r_{k+1}$. It is easy to verify that $g_{\mathbf{z},r_{k-1}}\leqslant f\leqslant g_{\mathbf{z},r_{k+2}}$. Our choice of m and the continuity condition (3.9) lead to $\int(g_{\mathbf{z},r_{k+2}}-g_{\mathbf{z},r_{k-1}})\,d\mu\leqslant\delta$.

Thus, Y is a one-sided (μ,δ)-cover of $\mathscr{K}_{\|\cdot\|}$. The cardinality of Y does not exceed

$$2qm^{q-1}\frac{m+3\kappa_1\kappa_2}{\kappa_1\kappa_2}\leqslant\frac{4q}{\kappa_1\kappa_2}m^q\leqslant\frac{4q}{\kappa_1\kappa_2}(7\kappa_1\kappa_2M_1)^q\delta^{-q/\gamma_1}.$$

Therefore,

$$H(\mathscr{K}_{\|\cdot\|},\mu,\delta)\leqslant\log\left\{\frac{4q}{\kappa_1\kappa_2}(7\kappa_1\kappa_2M_1)^q\right\}-(q/\gamma_1)\log\delta,$$

Along with Theorem 5.1, this leads to (5.18). □

Proof of Theorem 3.2. The theorem follows immediately from (2.26) and Proposition 5.4. □

The proof of Theorem 3.3 will follow from the following proposition.

Proposition 5.5. (a) *Let μ be a probability measure satisfying a continuity condition with parameters (M,γ). Let $R>0$ and $2M\tau_{q-1,\|\cdot\|_{q-1,2}}R^q\geqslant 1$. Let*

$$B=\log\{16(14\sqrt{q}\tau_{q-1,\|\cdot\|_{q-1,2}}MR^q)^{q+1}\}. \tag{5.20}$$

Then for integer $n\geqslant GB$,

$$\mathscr{E}_n(\mathscr{S}(R),\mu)\leqslant\varDelta_n((q+1)/\gamma,B). \tag{5.21}$$

(b) *Let μ be a regular measure with parameters (L,β,M,γ) satisfying* (3.14). *Let*

$$B=\log\{32(14\tau_{q-1,\|\cdot\|_{q-1,2}}Mq^{(q+1)/2}L^q2^{q\beta+1/\gamma})^{q+1}\}. \tag{5.22}$$

Then for integer $n\geqslant GB$,

$$\mathscr{E}_n(\mathscr{S}(\infty),\mu)\leqslant\varDelta_n((q+1)(q\beta+1/\gamma),B). \tag{5.23}$$

In order to prove this proposition, we first prove a simple lemma, estimating the volume of intersections of strips and spheres.

Lemma 5.3. *Let* $\mathbf{y}\in\mathbb{S}^{q-1}_{||\cdot||_2}$, $R>0$, $-R\leqslant a<b\leqslant R$. *Then*

$$\lambda_q(S(\mathbf{y},a,b)\cap B(||\cdot||_2,\mathbf{0},R))\leqslant\tau_{q-1,||\cdot||_{q-1,2}}R^{q-1}(b-a). \tag{5.24}$$

Proof. Since λ_q is rotation-invariant, we may assume that $\mathbf{y}=(0,\dots,0,1)$. Let $C(R,a,b)$ be the right cylinder with cross sections congruent to $B(||\cdot||_{q-1,2},\mathbf{0},R)$, base in the plane $x_{q+1}=a$ and top in the plane $x_{q+1}=b$. Then $S(\mathbf{y},a,b)\cap B(||\cdot||_{q,2},\mathbf{0},R)\subseteq C(R,a,b)$. Estimate (5.24) is now clear. □

Proof of Proposition 5.5. In this proof, we will denote $\mathbb{S}^{q-1}_{||\cdot||_2}$ by $\mathbb{S}^{q-1}$ and $\tau_{q-1,||\cdot||_{q-1,2}}$ by τ_{q-1}. Let $\delta\in(0,1]$. In view of the condition $2M\tau_{q-1}R^q\geqslant 1$, there exists an integer $m\geqslant 6\sqrt{q}$ in the interval

$$[12\sqrt{q}\tau_{q-1}MR^q\delta^{-1/\gamma},14\sqrt{q}\tau_{q-1}MR^q\delta^{-1/\gamma}].$$

As in the proof of Proposition 5.4, we find a set $\mathscr{C}$ consisting of $2qm^{q-1}$ points on $\mathbb{S}^{q-1}$ such that for any $\mathbf{y}\in\mathbb{S}^{q-1}$, there exists $\mathbf{z}\in\mathscr{C}$ with $||\mathbf{y}-\mathbf{z}||_2\leqslant 2\sqrt{q}/m$. Let $r_k=-R+2kR\sqrt{q}/m$, $-2\leqslant k\leqslant 2+m/\sqrt{q}$ (k integer). Let $g_{\mathbf{z},\ell,k}$ denote the characteristic function of $S(\mathbf{z},r_\ell,r_k)\cap B(||\cdot||_2,\mathbf{0},R)$, and $Y_\delta(R)$ be the set consisting of $\mathbb{1}$, $1-\mathbb{1}$, and these functions. Now, let f be the characteristic function of $S(\mathbf{y},a,b)\cap B(||\cdot||_2,\mathbf{0},R)$ for some $\mathbf{y}\in\mathbb{S}^{q-1}$, $[a,b]\subseteq\mathbb{R}$. We may assume that $[a,b]\subseteq[-R,R]$. We find a $\mathbf{z}\in\mathscr{C}$ with $||\mathbf{y}-\mathbf{z}||_2\leqslant 2\sqrt{q}/m$, and integers $\ell,k,0\leqslant\ell,k\leqslant m/\sqrt{q}$ such that $[r_{\ell+1},r_{k-1}]\subseteq[a,b]\subseteq[r_\ell,r_k]$. It is easy to verify that

$$g_{\mathbf{z},\ell+2,k-2}\leqslant f\leqslant g_{\mathbf{z},\ell-1,k+1}.$$

In view of Lemma 5.3, we verify that

$$\begin{aligned}&\int(g_{\mathbf{z},\ell-1,k+1}-g_{\mathbf{z},\ell+2,k-2})\,d\lambda_q\\&\leqslant\int(g_{\mathbf{z},\ell-1,\ell+2}+g_{\mathbf{z},k-2,k+1})\,d\lambda_q\\&\leqslant\frac{12\sqrt{q}\tau_{q-1}R^q}{m}.\end{aligned}$$

The continuity condition on μ and our choice of m now lead to the estimate

$$\int(g_{\mathbf{z},\ell-1,k+1}-g_{\mathbf{z},\ell+2,k-2})\,d\mu\leqslant\delta.$$

Thus, $Y_\delta(R)$ is a one-sided (μ,δ)-cover of $\mathscr{S}(R)$. Its cardinality does not exceed

$$\begin{aligned}|Y_\delta(R)|&\leqslant 2qm^{q-1}(5+m/\sqrt{q})^2+2\leqslant 8m^{q+1}\\&\leqslant 8(14\sqrt{q}\tau_{q-1}MR^q)^{q+1}\delta^{-(q+1)/\gamma}.\end{aligned} \tag{5.25}$$

Theorem 5.1 now leads to (5.21).

To prove part (b), we let $R = L(\delta/2)^{-\beta}$, h be the characteristic function of $\mathbb{R}^q \backslash [-R, R]^q$, and

$$Y := Y_{\delta/2}(\sqrt{q}R) \cup \{g + h : g \in Y_{\delta/2}(\sqrt{q}R)\}.$$

As in the proof of Proposition 5.2(b), Y is a one-sided (μ, δ)-cover of $\mathscr{S}(\infty)$, and $|Y| \leqslant 2|Y_{\delta/2}(\sqrt{q}R)|$. Estimate (5.25) with $\delta/2$ in place of δ and $\sqrt{q}L(\delta/2)^{-\beta}$ in place of R then leads to an estimate on $H(\mathscr{S}(\infty), \mu, \delta)$, which, along with Theorem 5.1 implies (5.23). $\square$

Proof of Theorem 3.3. The theorem follows immediately from (2.26) and Proposition 5.5. $\square$

Proof of Theorem 4.1. There is no loss of generality in assuming that ϕ is nondecreasing. Let

$$f(\mathbf{x}) = \int_{\mathbb{R}^q} \phi(\mathbf{x} \cdot \mathbf{y})\, d\mu(\mathbf{y}), \quad \mathbf{x} \in \mathbb{R}^q,$$

for a regular measure μ with parameters (L, β, M, γ) satisfying (3.14). Again, without loss of generality, we may assume that μ is a positive measure. In this proof, we will write $||\cdot||$ in place of $||\cdot||_2$. We observe that $f(\mathbf{0}) = \phi(0)$. Let $\mathbf{x} \in \mathbb{R}^q$, $\mathbf{x} \neq \mathbf{0}$, and $\mathbf{X} := \mathbf{x}/||\mathbf{x}||$. We note that for $\mathbf{y} \in \mathbb{R}^q$,

$$\phi(\mathbf{x} \cdot \mathbf{y}) = \int_{\mathbb{R}} \chi((-\infty, \mathbf{x} \cdot \mathbf{y}]; u)\, d\phi(u) = \int_{\mathbb{R}} \chi((-\infty, \mathbf{X} \cdot \mathbf{y}]; u/||\mathbf{x}||)\, d\phi(u).$$

Using Fubini's theorem, we obtain the representation

$$f(\mathbf{x}) = \int_{\mathbb{R}^q} \phi(\mathbf{x} \cdot \mathbf{y})\, d\mu(\mathbf{y}) = \int_{\mathbb{R}} \int_{\mathbb{R}^q} \chi(S(\mathbf{X}, u/||\mathbf{x}||, \infty); \mathbf{y})\, d\mu(\mathbf{y})\, d\phi(u). \quad (5.26)$$

In view of Proposition 5.5 (and its proof via Theorem 5.1), for $n \geqslant GB$, there exist points $\mathbf{y}_j \in \mathbb{R}^q$ such that

$$\left| \int_{\mathbb{R}^q} \chi(S(\mathbf{X}, u/||\mathbf{x}||, \infty); \mathbf{y})\, d\mu(\mathbf{y}) - \frac{1}{n} \sum_{j=1}^{n} \chi(S(\mathbf{X}, u/||\mathbf{x}||, \infty); \mathbf{y}_j) \right| \leqslant \Delta_n(\kappa, B), \quad (5.27)$$

where, in this proof only, $\kappa = (q+1)(q\beta + 1/\gamma)$. Now, we observe again that

$$\int_{\mathbb{R}} \chi(S(\mathbf{X}, u/||\mathbf{x}||, \infty); \mathbf{y}_j)\, d\phi(u) = \int_{\mathbb{R}} \chi((-\infty, \mathbf{x} \cdot \mathbf{y}_j]; u)\, d\phi(u) = \phi(\mathbf{x} \cdot \mathbf{y}_j).$$

Consequently, (5.26) and (5.27) lead to estimate (4.1) $\square$

Proof of Theorem 4.2. Without loss of generality, we assume that ϕ is nonincreasing. Let

$$f(\mathbf{x}) = \int_{\mathbb{R}^q} \phi(||\mathbf{x}-\mathbf{y}||)\, d\mu(\mathbf{y})$$

for a regular measure μ with parameters (L, β, M, γ), where we may assume without loss of generality that μ is a positive measure. Let $\mathbf{x} \in \mathbb{R}^q$. Writing $dv(u) = -d\phi(u)$, we note that

$$\begin{aligned} f(\mathbf{x}) &= -\int_{\mathbb{R}^q} \int_0^\infty \chi([||\mathbf{x}-\mathbf{y}||, \infty); u)\, d\phi(u)\, d\mu(\mathbf{y}) \\ &= \int_0^\infty \int_{\mathbb{R}^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\, d\mu(\mathbf{y})\, dv(u). \end{aligned} \tag{5.28}$$

Using (4.2) and the decay condition on μ, we derive that

$$\begin{aligned} &\left| f(\mathbf{x}) - \int_0^{R_1} \int_{[-R_1,R_1]^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\, d\mu(\mathbf{y})\, dv(u) \right| \\ &\quad \leqslant \int_{(u,\mathbf{y}) \in [0,\infty] \times \mathbb{R}^q \setminus [0,R_1] \times [-R_1,R_1]^q} d\mu(\mathbf{y})\, dv(u) \leqslant \sqrt{\log n/n} \leqslant 1/2. \end{aligned} \tag{5.29}$$

First, we consider the case when $||\mathbf{x}||_\infty \leqslant R$. Writing $I = \int_{[-R_1,R_1]^q} d\mu(\mathbf{y})$ and $I_1 = \int_0^{R_1} dv(u)$, we see that for any measurable function $g : [0, \infty) \to [-1, 1]$,

$$\left| \int_0^\infty g(u)\, dv(u) - \frac{1}{I_1} \int_0^{R_1} g(u)\, dv(u) \right| \leqslant 3\sqrt{\log n/n} \tag{5.30}$$

and

$$\begin{aligned} &\left| \int_0^{R_1} \int_{[-R_1,R_1]^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\, d\mu(\mathbf{y})\, d\phi(u) \right. \\ &\quad \left. - \frac{1}{II_1} \int_0^{R_1} \int_{[-R_1,R_1]^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\, d\mu(\mathbf{y})\, dv(u) \right| \\ &\quad \leqslant 3\sqrt{\log n/n}. \end{aligned}$$

Therefore,

$$\left| f(\mathbf{x}) - \frac{1}{II_1} \int_0^{R_1} \int_{[-R_1,R_1]^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\, d\mu(\mathbf{y})\, dv(u) \right| \leqslant 4\sqrt{\log n/n}. \tag{5.31}$$

The measure $\frac{1}{I} d\mu(\mathbf{y})$, supported on $[-R_1, R_1]^q$ satisfies the continuity condition with parameters $(2^{1/\gamma} M, \gamma)$. Since condition (3.5) is satisfied, we may apply Proposition 5.3 (and its proof via Theorem 5.4) to obtain points $\mathbf{y}_j \in [-R_1, R_1]^q$ such that for

$u \in [0, R_1]$,

$$\left| \frac{1}{I} \int_{[-R_1, R_1]^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\ d\mu(\mathbf{y}) - \frac{1}{n} \sum_{j=1}^{n} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y}_j) \right| \leqslant \Delta_n(\kappa, B),$$

where, in this proof only, $\kappa = 3q\gamma^{-1}$. Consequently,

$$\left| \frac{1}{II_1} \int_0^{R_1} \int_{[-R_1, R_1]^q} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\ d\mu(\mathbf{y})\ dv(u) \right.$$
$$\left. - \frac{1}{nI_1} \sum_{j=1}^{n} \int_0^{R_1} \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y}_j)\ dv(u) \right|$$
$$\leqslant \Delta_n(\kappa, B).$$

In view of (5.30) and (5.31) this leads to

$$\left| f(\mathbf{x}) - \frac{1}{n} \sum_{j=1}^{n} \phi(||\mathbf{x} - \mathbf{y}_j||) \right| \leqslant \Delta_n(\kappa, B) + 7\sqrt{\frac{\log n}{n}}.$$

This proves (4.3) in the case when $||\mathbf{x}||_\infty \leqslant R$.

Next, let $||\mathbf{x}||_\infty > R$. Then $||\mathbf{x}|| > (1 + \kappa_2)R_1$, and for $(u, \mathbf{y}) \in [0, R_1] \times [-R_1, R_1]^q$, we obtain

$$||\mathbf{x} - \mathbf{y}|| \geqslant ||\mathbf{x}|| - ||\mathbf{y}|| > (1 + \kappa_2)R_1 - \kappa_2 ||\mathbf{y}||_\infty \geqslant R_1 \geqslant u;$$

i.e., $\chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y}) = 0$. Therefore, for all $\mathbf{y} \in [-R_1, R_1]^q$,

$$|\phi(||\mathbf{x} - \mathbf{y}||)| = \left| \int_0^\infty \chi(B(||\cdot||, \mathbf{x}, u); \mathbf{y})\ dv(u) \right|$$
$$\leqslant \int_{R_1}^\infty dv(u) \leqslant \sqrt{\frac{\log n}{n}},$$

and in particular,

$$\left| \frac{1}{n} \sum_{j=1}^{n} \phi(||\mathbf{x} - \mathbf{y}_j||) \right| \leqslant \sqrt{\frac{\log n}{n}}.$$

Moreover, (5.29) implies that $|f(\mathbf{x})| \leqslant \sqrt{\log n / n}$. Hence,

$$\left| f(\mathbf{x}) - \frac{1}{n} \sum_{j=1}^{n} \phi(||\mathbf{x} - \mathbf{y}_j||) \right| \leqslant 2\sqrt{\frac{\log n}{n}}. \qquad \square$$

Finally, we prove the remaining assertion of this paper, Proposition 2.1.

Proof of Proposition 2.1. In this proof only, we will denote the surface area of the Euclidean unit sphere embedded in $\mathbb{R}^q$ by

$$\omega_{q-1} := \frac{2\pi^{q/2}}{\Gamma(q/2)}. \tag{5.32}$$

Passing to spherical coordinates, we see that

$$\lambda_{\exp,\alpha}^{-1} = \int_{\mathbb{R}^q} \exp(-||\mathbf{x}||_2^\alpha)\, d\mathbf{x} = \omega_{q-1} \int_0^\infty r^{q-1} e^{-r^\alpha}\, dr$$
$$= \frac{\omega_{q-1}}{\alpha} \int_0^\infty t^{q/\alpha-1} e^{-t} dt = \frac{\omega_{q-1}\Gamma(q/\alpha)}{\alpha}. \tag{5.33}$$

Since $\exp(-||\mathbf{x}||_2^\alpha) \leqslant 1$, the assertion about the continuity condition is now clear. To prove the assertion regarding the decay condition, let (in this proof only) $s := (q-\alpha)/\alpha$, and $s! := \Gamma(q/\alpha)$. Passing to the spherical coordinates again, a little calculation as above leads to

$$\mu_{\exp}(\alpha; \mathbb{R}^q \backslash [-R, R]^q) \leqslant \frac{1}{s!} \int_{R^\alpha}^\infty t^s e^{-t}\, dt, \quad R > 0. \tag{5.34}$$

Now, an integration by parts shows that if $R \geqslant 1$,

$$\int_{R^\alpha}^\infty t^s e^{-t}\, dt = R^{\alpha s} \exp(-R^\alpha) + s \int_{R^\alpha}^\infty t^{s-1} e^{-t}\, dt \leqslant R^{\alpha|s|} \exp(-R^\alpha)$$
$$+ \frac{|s|}{R^\alpha} \int_{R^\alpha}^\infty t^s e^{-t}\, dt.$$

It follows that if $R^\alpha > \max(2|s|, 1)$, then

$$\mu_{\exp}(\alpha; \mathbb{R}^q \backslash [-R, R]^q) \leqslant \frac{1}{s!} \int_{R^\alpha}^\infty t^s e^{-t}\, dt \leqslant \frac{2}{s!} R^{\alpha|s|} \exp(-R^\alpha). \tag{5.35}$$

Now, let $\delta > 0$, and in this proof only, let $\varepsilon = (s!/2)\delta$. Using elementary calculus, we verify that $x - a\log x \geqslant a \log(e/a)$ for every $x > 0$ and $a > 0$. Therefore, choosing

$$\frac{\log A}{\alpha\beta} := (|s| + 1/(\alpha\beta)) \log\left(\frac{|s| + 1/(\alpha\beta)}{e}\right), \quad x = A\varepsilon^{-\alpha\beta}, \quad a = |s| + 1/\alpha\beta,$$

we conclude that $(A\varepsilon^{-\alpha\beta})^{|s|} \exp(-A\varepsilon^{-\alpha\beta}) \leqslant \varepsilon$. Therefore, with $R^\alpha \geqslant \max(1, 2|s|, A\varepsilon^{-\alpha\beta})$, we see from (5.35) that $\mu_{\exp}(\alpha; \mathbb{R}^q \backslash [-R, R]^q) \leqslant \delta$. This completes the proof of part (a).

For the proof of part (b), we recall the identity (cf. [3, Chapter V, Example 2.12])

$$\int_0^\infty \frac{x^{t-1}}{1+x} = \frac{\pi}{\sin \pi t}, \quad 0 < t < 1.$$

The remainder of the proof of part (b) using spherical coordinates is very elementary, and is omitted. □

Acknowledgments

The paper was originally intended to be a joint work with Professor Steven Damelin. In particular, he activated me to pursue this research, which had lain dormant in my mind for several years. In a series of emails, especially related to Theorem 5.1, Propositions 5.2 and 5.4, he asked many questions, thus highlighting what parts of the first draft might be obscure to some readers. I am thankful to him for his efforts. I am grateful to Professor Grzegorz Wasilkowski for a careful reading of another draft and for making many useful suggestions, resulting in a substantial improvement of the results in that draft. Finally, I am grateful both to Professor Grzegorz Wasilkowski and Professor Henryk Woźniakowski for their keen interest in this work.

References

[1] A.R. Barron, Neural net approximation, in: K.S. Narendra (Ed.), Proceedings of 7th Yale Workshop on Adaptive and Learning Systems, Yale University Press, New Haven, Connecticut, 1992, pp. 69–72.
[2] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inform. Theory 39 (1993) 930–945.
[3] J.B. Conway, Functions of One Complex Variable, Springer, New York, 1973.
[4] S. Heinrich, E. Novak, G. W. Wasilkowski, H. Woźniakowski, The inverse of the star-discrepancy depends linearly on the dimension, Acta Arith. XCVI.3 (2001) 279–302.
[5] F.J. Hickernell, I.H. Sloan, G.W. Wasilkowski, On tractability of weighted integration over bounded and unbounded regions in R^s, Math. Comput., accepted for publication.
[6] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1999.
[7] V. Kurkova, M. Sanguineti, Bounds on rates of variable basis and neural network approximation, IEEE Trans. Inform. Theory 47 (2001) 2659–2665.
[8] V. Kurkova, M. Sanguineti, Comparison of worst case errors in linear and neural network approximation, IEEE Trans. Inform. Theory 48 (2002) 264–275.
[9] Y. Makovoz, Random approximants and neural networks, J. Approx. Theory 85 (1996) 98–109.
[10] Y. Makovoz, Uniform approximation by neural networks, J. Approx. Theory 95 (1998) 215–228.
[11] H.N. Mhaskar, Approximation theory and neural networks, in: P.K. Jain, M. Krishnan, H.N. Mhaskar J. Prestin, D. Singh (Eds.), Wavelet Analysis and Applications, Proceedings of the International Workshop in Delhi, 1999, Narosa Publishing, New Delhi, India, 2001, pp. 247–289.
[12] H.N. Mhaskar, C.A. Micchelli, Dimension-independent bounds on approximation by neural networks, IBM J. Res. Development 38 (1994) 277–284.
[13] E. Novak, H. Woźniakowski, When are integration and discrepancy tractable?, in: R. DeVore, A. Iserles, E. Süli (Eds.), Foundations of Computational Mathematics, Cambridge University Press, Cambridge, 2001, pp. 211–266.
[14] E. Novak, H. Woźniakowski, Intractability results for integration and discrepancy, J. Complexity 17 (2001) 388–441.
[15] S.H. Paskov, New methodologies for evaluating derivatives, in: S. Pliska, M. Dempster (Eds.), Mathematics of Derivative Securities, Cambridge University Press, Cambridge, 1997, pp. 545–582.
[16] S.H. Paskov, J.F. Traub, Faster valuation of financial securities, J. Portfolio Management 22 (1995) 113–120.
[17] D. Pollard, Convergence of Stochastic Processes, Springer, New York, 1984.
[18] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1974.
[19] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, with Applications to Statistics, Springer, New York, 1996.
[20] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.