METRIC ENTROPY OF HIGH DIMENSIONAL DISTRIBUTIONS

RON BLEI, FUCHANG GAO, AND WENBO V. LI

(Communicated by Richard C. Bradley)

ABSTRACT. Let $\mathcal{F}_d$ be the collection of all d-dimensional probability distribution functions on $[0,1]^d$, $d \geq 2$. The metric entropy of $\mathcal{F}_d$ under the $L_2([0,1]^d)$ norm is studied. The exact rate is obtained for $d = 1, 2$ and bounds are given for $d > 3$. Connections with small deviation probability for Brownian sheets under the sup-norm are established.

1. INTRODUCTION

Metric entropy is an important tool that has been widely used in many areas. Roughly put, it is a quantification or measurement of compactness. Given a set T in a Banach space $(E, \|\cdot\|)$ and given a small $\varepsilon > 0$, the metric entropy of T can be defined as $\log N(T, \|\cdot\|, \varepsilon)$, where $N(T, \|\cdot\|, \varepsilon)$ is the minimum covering number

$$N(T, \|\cdot\|, \varepsilon) := \inf\left\{ n : \exists t_1, t_2, \ldots, t_n \in T, \text{ s.t. } T \subset \bigcup_{k=1}^{n} B(t_k, \varepsilon) \right\}$$

and $B(x, \varepsilon)$ is the open ε-ball in E centered at x. When there is no confusion regarding the norm, we write $\log N(T, \varepsilon)$. The precise definition of the metric entropy was given first by A. N. Kolmogorov, and its various asymptotic behaviors, as $\varepsilon \to 0^+$, have subsequently been studied and applied in approximation theory, geometric functional analysis, probability theory, and complexity theory; e.g., see the books by Kolmogorov and Tihomirov [8], Lorentz [12], Carl and Stephani [3], and Edmunds and Triebel [5].

In this paper we consider the case where T is the collection of probability distributions on $[0,1]^d$, denoted by $\mathcal{F}_d$, and we study its metric entropy estimate under L^2-norm. Because absolutely continuous functions are dense in $\mathcal{F}_d$ under L^2-norm, we assume that $\mathcal{F}_d$ consists only of distributions that have a density. This problem arises naturally in non-parametric estimation in statistics; see van der Vaart and Wellner [14].

The main contributions of this paper are the following estimates and the establishment of very useful connections between the metric entropy of $\mathcal{F}_d$ under L^2-norm and the small deviation probability for Brownian sheets under the sup-norm.

Theorem 1.1. *For $d \geq 2$ there exist constants $c_1, c_2 > 0$ depending only on d such that for $0 < \varepsilon < e^{-1}$,*

$$c_1 \varepsilon^{-1} [\log(1/\varepsilon)]^{d-1} \leq \log N(\mathcal{F}_d, \|\cdot\|_2, \varepsilon) \leq c_2 \varepsilon^{-1} [\log(1/\varepsilon)]^{d-1/2}.$$

Moreover, for $d = 2$,

$$c_1 \varepsilon^{-1} [\log(1/\varepsilon)]^{3/2} \leq \log N(\mathcal{F}_2, \|\cdot\|_2, \varepsilon) \leq c_2 \varepsilon^{-1} [\log(1/\varepsilon)]^{3/2}.$$

The result for the case $d = 1$ is known. Specifically, van der Vaart and Wellner [14] contains (on page 159) a proof of the upper bound estimate of the covering number. We give a shorter proof in the next section.

We underscore a fundamental duality relation for metric entropies, which changes the focus from one norm to another. Indeed, it is a crucial tool for our proof of Theorem 1.1. Let T° be the polar body of T, and let B^* be the unit ball of the dual space E^* of E. If one of the norms is the L^2-norm, then for small $\varepsilon > 0$,

$$\log N(T, \|\cdot\|, \varepsilon) \approx \log N(B^*, \|\cdot\|_{T^\circ}, c\varepsilon)$$

up to a multiplicative constant. Here $a \approx b$ means $c_1 \leq a/b \leq c_2$ for some constants $0 < c_1 < c_2 < \infty$. The duality relation in the general case has been a conjecture for more than 30 years; see Artstein et al. [1] and [2].

The remaining sections Sare organized as follows. Section 2 deals with the case $d = 1$ and points to the main difficulties in the higher dimensional case. Section 3 presents the useful connection with small ball probabilities, which is based on the duality relation (stated above) and fundamental links between metric entropy and small ball probabilities for Gaussian measures (discovered in Kuelbs and Li [9] and completed in Li and Linde [10]). In fact, the main significance of this paper is the discovery of this connection and using it to obtain previously unknown estimates on metric entropy of distribution functions. Whereas the upper and lower bound estimates in Theorem 1.1 can both be obtained from known results on small ball estimates, we present in Section 4 a direct and new argument establishing the lower bound. This provides the simplest known proof of the corresponding result for small ball estimates in the two-dimensional case. Two different (and more difficult) proofs in the small ball setting are presented in Talagrand [13], Gao and Li [7]. For a survey of small ball probabilities and various applications, see Li and Shao [11].

2. The case $d = 1$

The upper bound in the following result can be found in [14]. Our proof below is somewhat simpler. The lower bound may also be known, but we were unable to locate a reference.

Proposition 2.1. *There exist constants c_1 and c_2 such that*

$$c_1 \varepsilon^{-1} \leq \log N(\mathcal{F}_1, \|\cdot\|_2, \varepsilon) \leq c_2 \varepsilon^{-1}.$$

Proof. For convenience, we assume $1/\varepsilon = n$ is an even integer. To obtain the lower bound, we construct random functions on $[0, 1]$ by

$$f_\eta(x) = \begin{cases} 0, & x = 0, \\ (k - 1/2)\varepsilon + \eta_k \varepsilon/2, & x \in [(k-1)\varepsilon, k\varepsilon), \quad 1 \leq k \leq n, x \neq 0, x \neq 1, \\ 1, & x = 1, \end{cases}$$

where $\eta_k \in \{-1, 1\}$. Let $\mathcal{U}$ be the collection of all such functions. Clearly $\mathcal{U}$ contains 2^n distinct functions, each of which is a probability distribution function on $[0, 1]$. Given $f_\eta \in \mathcal{U}$ and $f_{\eta'} \in \mathcal{U}$, define the Hamming distance

$$d_h(f_\eta, f_{\eta'}) = \text{card}\{k : \eta_k \neq \eta'_k, 1 \leq k \leq n\}.$$

Clearly, for each $f \in \mathcal{U}$, the set $\{g \in \mathcal{U} : d_h(f, g) < \lceil n/10 \rceil\}$ contains no more than $\sum_{k=0}^{\lceil n/10 \rceil} \frac{n}{k} \leq 2^{n/2}$ functions. Thus, we can find $N = 2^n/2^{n/2} = 2^{n/2}$ functions $f_1, f_2, ..., f_N \in \mathcal{U}$, such that $d_h(f_i, f_j) \geq \lceil n/10 \rceil$ for all $i \neq j$, $1 \leq i, j \leq N$. Thus, $|f_i(x) - f_j(x)| = \varepsilon$ on at least $\lceil n/10 \rceil$ intervals $((k-1)\varepsilon, k\varepsilon)$ of length ε. Hence $\|f_i - f_j\|_2 \geq \varepsilon/\sqrt{10}$. Therefore $N(\mathcal{F}_1, \|\cdot\|_2, \varepsilon/\sqrt{10}) \geq 2^{n/2}$, which implies the lower bound in the proposition.

To obtain the upper bound, let $\mathcal{F}_{[0,a]}$ be the collection of probability distribution functions on $[0, a]$. Each $f \in \mathcal{F}_{[0,1]}$ can be approximated by step functions from below and from above, and, therefore, f can be assumed always to take values in $\{0, \varepsilon, 2\varepsilon, ..., n\varepsilon\}$. Consider the intervals $[(k-1)\varepsilon/2, k\varepsilon/2)$ that are contained in $[0, 1]$. Because f is nondecreasing and takes value in $\{0, \varepsilon, 2\varepsilon, ..., n\varepsilon\}$, f is constant on at least a half of these $2n$ intervals. Choose n such intervals I_1, I_2, ..., I_n, ordered from left to right. Let $a_1 \leq a_2 \leq \cdots \leq a_n$ be the respective values of f on these intervals. Define $g(0) = 0$ and

$$g(x) = \sum_{i=1}^{n} a_i 1_{I_i}(x)$$

for $x \in (0, 1]$. Because there are $\binom{2n}{n}$ choices of $\{I_1, I_2, ..., I_n\}$ and because for each choice of $\{I_1, I_2, ..., I_n\}$ there are no more than $\binom{n+n}{n}$ choices of the values $a_1 \leq a_2 \leq \cdots \leq a_n$, the total number of choices of g is no more than $\binom{2n}{n}^2$. Once g is fixed, consider the function $f - g$. By deleting $I_1, ..., I_n$ from $[0, 1]$ and shifting the remaining intervals to the left, $f - g$ corresponds to a probability distribution function $\xi \in \mathcal{F}_{[0,1/2]}$. Write $T_g(f - g) = \xi$, where T_g is the shifting operator depending on g. Let S be a 10ε-net of $\mathcal{F}_{[0,1/2]}$ with cardinality $N(\mathcal{F}_{[0,1/2]}, \|\cdot\|, 10\varepsilon)$. Then, there exists a $\mu \in S$, such that $\|\xi - \mu\|_2 \leq 10\varepsilon$. Hence, $\|f - (g + T_g^{-1}\mu)\|_2 = \|\xi - \mu\|_2 \leq 10\varepsilon$. Because there are no more than $N(\mathcal{F}_{[0,1/2]}, \|\cdot\|_2, 10\varepsilon) \cdot \binom{2n}{n}^2$ realizations of $g + T_g^{-1}\mu$, we obtain

$$(2.1) \qquad N(\mathcal{F}_{[0,1]}, \|\cdot\|_2, 12\varepsilon) \leq \binom{2n}{n}^2 \cdot N(\mathcal{F}_{[0,1/2]}, \|\cdot\|_2, 10\varepsilon).$$

Note that a set is a 10ε-net of $\mathcal{F}_{[0,1/2]}$ if and only if its dilation $\{f(t/2) : f(t) \in U\}$ is a $10\sqrt{2}\varepsilon$-net of $\mathcal{F}_{[0,1]}$. Thus, $N(\mathcal{F}_{[0,1/2]}, \|\cdot\|_2, 10\varepsilon) = N(\mathcal{F}_{[0,1]}, \|\cdot\|_2, 10\sqrt{2}\varepsilon)$. Putting this in (2.1) and using the fact that $\binom{2n}{n}^2 \leq 2^{4n} = 2^{4/\varepsilon}$, we obtain

$$N(\mathcal{F}_{[0,1]}, \|\cdot\|_2, 12\varepsilon) \leq 2^{4/\varepsilon} \cdot N(\mathcal{F}_{[0,1]}, \|\cdot\|_2, 10\sqrt{2}\varepsilon).$$

By iteration, we obtain the upper bound. $\qquad \square$

Remark 2.2. The preceding proof shows that we can require $g + T^{-1}\mu \geq f$ or $g + T^{-1}\mu \leq f$. Thus, the upper estimate holds also for bracketing entropy, as given in [14].

Remark 2.3. In the proof of the lower bound, we used the fact that $\mathcal{F}_1$ consists essentially of all increasing functions $f(x)$ on $[0, 1]$ with $f(0) = 0$ and $f(1) = 1$. An

obvious difficulty in the higher dimensional case is the lack of a simple analogous characterization of $\mathcal{F}_d$. Thus, a different method is needed to construct a large, well-separated set of functions in $\mathcal{F}_d$.

3. Connection with small ball probability of Brownian sheets

First we note
$$N(\mathcal{F}_d, \|\cdot\|_2, \varepsilon) = N(\mathcal{E}_d, \|\cdot\|_2, \varepsilon),$$
where
$$\mathcal{E}_d = \left\{ \int_{t_1}^1 \int_{t_2}^1 \cdots \int_{t_d}^1 d\mu : \mu \text{ is a probability measure} \right\}.$$
Indeed, a finite set $\{f_i\}$ is an ε-net of $\mathcal{F}_d$ if and only if the set $\{g_i\}$ is an ε-net of $\mathcal{E}_d$, where
$$g_i(t_1, t_2, ..., t_d) = f_i(1 - t_1, 1 - t_2, ..., 1 - t_d).$$

Let $\{\phi_{k_1} \otimes \cdots \otimes \phi_{k_d} : k_i \in \mathbb{N}, 1 \le i \le d\}$ be an orthonormal basis of $L^2[0,1]^d$. For $f, g \in \mathcal{E}_d$, let $a = (a_{k_1,...,k_d})$ and $b = (b_{k_1,...,k_d})$ be the Fourier coefficients of f and g, respectively, relative to $\{\phi_{k_1} \otimes \cdots \otimes \phi_{k_d}\}$. By Parseval's identity, $\|f - g\|_2 = \|a - b\|_{l^2}$. Thus,
$$N(\mathcal{F}_d, \|\cdot\|_2, \varepsilon) = N(\mathcal{E}_d, \|\cdot\|_2, \varepsilon) = N(S, \|\cdot\|_{l^2}, \varepsilon),$$
where
$$S = \left\{ (a_{k_1,...,k_d}) = \langle g, \phi_{k_1} \otimes \cdots \otimes \phi_{k_d} \rangle : g \in \mathcal{E}_d \right\}.$$
If μ is the probability measure corresponding to $g \in \mathcal{E}_d$, then using integration by parts, we obtain
$$\langle g, \phi_{k_1} \otimes \cdots \otimes \phi_{k_d} \rangle = \left\langle \mu, \int_0^{t_1} \cdots \int_0^{t_d} \phi_{k_1} \otimes \cdots \otimes \phi_{k_d} \right\rangle.$$
In other words,
$$S = \left\{ (a_{k_1,...,k_d}) : a_{k_1,...,k_d} = \left\langle \mu, \int_0^{t_1} \cdots \int_0^{t_d} \phi_{k_1} \otimes \cdots \otimes \phi_{k_d} \right\rangle, \|\mu\|_1 = 1; \mu \ge 0 \right\}.$$
Let
$$T = \left\{ (a_{k_1,...,k_d}) : a_{k_1,...,k_d} = \left\langle \mu, \int_0^{t_1} \cdots \int_0^{t_d} \phi_{k_1} \otimes \cdots \otimes \phi_{k_d} \right\rangle, \|\mu\|_1 \le 1 \right\}.$$
Then, clearly $T \supset S$. Thus,
$$\log N(T, \|\cdot\|_{l^2}, \varepsilon) \ge \log N(S, \|\cdot\|_{l^2}, \varepsilon) = \log N(\mathcal{F}_d, \|\cdot\|_2, \varepsilon).$$
On the other hand, for any μ with $\|\mu\|_1 \le 1$, there are probability measures μ_1 and μ_2, such that $\mu = \mu_1 - \mu_2$. Thus, $T \subset S - S$, which implies that
$$\log N(T, \|\cdot\|_{l^2}, \varepsilon) \le 2 \log N(S, \|\cdot\|_{l^2}, \varepsilon/2) = 2 \log N(\mathcal{F}_d, \|\cdot\|_2, \varepsilon/2).$$
By the duality relation stated in the introduction, we have
$$\log N(T, \|\cdot\|_{l^2}, \varepsilon) \approx c_1 \log N(B_{l^2}, T^\circ, c_2 \varepsilon),$$
where B_{l^2} is the unit ball of l^2 and T° is the polar body of T. Thus,
$$\log N(\mathcal{F}_d, \|\cdot\|_2, \varepsilon) \approx c_1 \log N(B_{l^2}, T^\circ, c_2 \varepsilon).$$
Note that T is convex; in fact, T is precisely the absolute convex hull of the set
$$\left\{ (a_{k_1,...,k_d}) : a_{k_1,...,k_d} = \int_0^{t_1} \cdots \int_0^{t_d} \phi_{k_1} \otimes \cdots \otimes \phi_{k_d}, (t_1, ..., t_d) \in [0,1]^d \right\}$$

(cf. Gao [6]). This observation leads to the connection with small ball probability of Brownian sheets, to which we now turn our attention.

A d-dimensional Brownian sheet $B_d(t, \cdot)$, $t \in [0,1]^d$, $d \geq 2$, is a Gaussian process on $[0,1]^d$ with covariance kernel

$$K(t,s) = \prod_{i=1}^{d} \min(t_i, s_i), \quad t = (t_1, t_2, ..., t_d), \quad s = (s_1, s_2, ..., s_d).$$

It is sometimes called the d-fold tensor product of standard Brownian motion.

It is easier to deal with its canonical series expansion

$$\begin{aligned} B_d(t, \cdot) &= \sum_{k_1,...,k_d} \xi_{k_1,...,k_d}(\cdot) T_d(\phi_{k_1} \otimes \cdots \otimes \phi_{k_d}) \\ &= \sum_{k_1,...,k_d} \xi_{k_1,...,k_d}(\cdot) \int_0^{t_1} \cdots \int_0^{t_d} \phi_{k_1} \otimes \cdots \otimes \phi_{k_d}, \end{aligned}$$

where T_d is the generating compact operator of $B_d(t)$ defined by

$$(3.1) \qquad T_d f(x_1, ..., x_d) = \int_0^{x_1} \cdots \int_0^{x_d} f(t_1, ..., t_d) dt_1 \cdots dt_d,$$

$\{\phi_{k_1} \otimes \cdots \otimes \phi_{k_d}\}$ is any complete orthonormal system in $L^2([0,1]^d)$, and $\xi_{k_1,...,k_d}(\cdot)$ are i.i.d. standard normal random variables. Now we observe that

$$\sup_{t \in [0,1]^d} |B_d(t, \omega)| \leq \varepsilon \Leftrightarrow (\xi_{k_1,...,k_d}(\omega)) \in \varepsilon T^{\circ}.$$

Therefore,

$$\mathbb{P}\left(\sup_{t \in [0,1]^d} |B_d(t, \cdot)| \leq \varepsilon\right) = \gamma(\varepsilon T^{\circ}),$$

where γ is the standard Gaussian measure.

The remaining part of this section is to relate $\gamma(\varepsilon T^{\circ})$ with $N(B_{l^2}, T^{\circ}, \varepsilon)$. Roughly put, the smaller the value of $\gamma(\varepsilon T^{\circ})$, the larger the covering number $N(B_{l^2}, T^{\circ}, \varepsilon)$. The remarkable discovery in [9] (completed in [10]) is that there is a tight connection between these two quantities: upper (lower) bound on one implies lower (upper) bound for the other. In particular, the statement is

$$\log \gamma(\varepsilon T^{\circ}) \approx -\varepsilon^{-2} [\log(1/\varepsilon)]^{\beta}$$

if and only if

$$\log N(B_{l^2}, T^{\circ}, \varepsilon) \approx \varepsilon^{-1} (\log 1/\varepsilon)^{\beta/2}.$$

However, it is also proved in Dunker et al. [4] that

$$\log \mathbb{P}\left(\sup_{t \in [0,1]^d} |B_d(t, \cdot)| \leq \varepsilon\right) \geq -c\varepsilon^{-2} [\log(1/\varepsilon)]^{2d-1}.$$

Therefore, the upper bound of Theorem 1.1 follows from the one-sided result in [10]. The lower bound also follows from the best known results on small ball probability; see [7] for details and history, and see also the discussion in the next section.

4. LOWER BOUND

As noted above, we could just as well have used the known upper bound for small ball probability of Brownian sheets to obtain the lower bound for entropy of distributions. However, the proof given below is new and different. Indeed, in view of the equivalence given above, it gives the simplest known proof regarding the upper bound of small deviation of Brownian sheet in the two-dimensional case. The key to the proof is the idea of random Riesz products and random sums, which we now proceed to explain.

Given $m \in \mathbb{N}$ (where $\mathbb{N}$ denotes the set of positive integers) and $1 \le i \le 2^m$, we let $\{\psi_{m,i}\}$ be the orthogonal, rescaled and shifted Haar functions defined by

$$\psi_{m,i}(t) = \begin{cases} 1, & t \in [(i-1)2^{-m}, (i-3/4)2^{-m}) \cup [(i-1/4)2^{-m}, i2^{-m}), \\ -1, & [(i-3/4)2^{-m}, (i-1/4)2^{-m}), \\ 0, & \text{otherwise.} \end{cases}$$

For a fixed even integer $n > 2d$ and for $1 \le k \le n$, we denote

$$\mathcal{M}_n := \{(m_1, m_2, ..., m_d) \in \mathbb{N}^d : m_1 + m_2 + \cdots + m_d = n, m_i \text{ is even}, 1 \le i \le d\}.$$

It is easy to check that $|\mathcal{M}_n| = \binom{n/2-1}{d-1} \sim (n/2)^{d-1}/(d-1)!$. For $M \in \mathcal{M}_n$, denote

$$\mathcal{I}_M = \{(i_1, ..., i_d) \in \mathbb{N}^d : i_j \le 2^{m_j}, 1 \le j \le d\},$$

and for $I = (i_1, i_2, ..., i_d) \in \mathcal{I}_M$, denote

(4.1)
$$\Psi_{M,I} = \psi_{m_1,i_1} \otimes \psi_{m_2,i_2} \otimes \cdots \otimes \psi_{m_d,i_d}.$$

For each $\eta = (\eta_{M,I})$, where the $\eta_{M,I}$ ($I \in \mathcal{I}_M$, $M \in \mathcal{M}_n$) are independent Bernoulli random variables (i.e., $\mathbb{P}(\eta_{M,I} = \pm 1) = 1/2$), we consider the random function

$$F_M(\cdot, \eta) = \sum_{I \in \mathcal{I}_M} \eta_{M,I} \Psi_{M,I}(\cdot)$$

and then define the random functions

$$R_n(\cdot, \eta) = \prod_{M \in \mathcal{M}_n} (1 + F_M(\cdot, \eta)) \quad \text{if } d = 2,$$

$$R_n(\cdot, \eta) = \frac{1}{\sqrt{|\mathcal{M}_n|}} \sum_{M \in \mathcal{M}_n} F_M(\cdot, \eta) \quad \text{if } d > 2.$$

Note that the F_M are orthogonal functions, with $\|F_M\|_2 = 1$, and therefore, for $d > 2$, $\|R_n\|_1 \le \|R_n\|_2 = 1$. Also note that $|F_M| = 1$. It is easy to check that for $d = 2$, we have $R_n \ge 0$ and $\|R_n(\cdot, \eta)\|_1 = 1$ for all η. Clearly, there are $2^{|\mathcal{M}_n|2^n}$ realizations of R_n. Between any two realizations $R_n(\cdot, \eta^{(1)})$ and $R_n(\cdot, \eta^{(2)})$, we consider the Hamming distance

$$d_h(R_n(\cdot, \eta^{(1)}), R_n(\cdot, \eta^{(2)})) = \text{card}\left\{(M, I) : I \in \mathcal{I}_M, M \in \mathcal{M}_n, \eta_{M,I}^{(1)} \ne \eta_{M,I}^{(2)}\right\}.$$

By a combinatorial argument similar to the one given in Section 2, we can find at least $N \approx \exp(c_1 n^{d-1} 2^n)$ realizations $R_n(\cdot, \eta^{(i)})$, $i \le N$, such that

$$d_h(R_n(\cdot, \eta^{(i)}), R_n(\cdot, \eta^{(j)})) \ge c_2 n^{d-1} 2^n$$

for all $i \ne j$. For each of such $R_n(\cdot, \eta^{(i)})$, $1 \le i \le N$, we define

$$f_i(x_1, x_2, ..., x_d) = \int_0^{x_1} \cdots \int_0^{x_d} R_n(\cdot, \eta^{(i)})(t_1, ..., t_d) dt_1 \cdots dt_d.$$

Next we show that there exists a constant c_3 such that for all $i \neq j$,

(4.2)
$$\|f_i - f_j\|_2 \geq \begin{cases} c_3\sqrt{n}2^{-n}, & d = 2, \\ c_3 2^{-n}, & d > 2. \end{cases}$$

By Bessel's inequality,

$$\|f_i - f_j\|_2^2 \geq \sum_{M \in \mathcal{M}_n} \sum_{I \in \mathcal{I}_M} |\langle f_i - f_j, H_{M,I}\rangle|^2,$$

where the $H_{M,I}$ are orthonormal Haar functions; i.e., $H_{M,I} = h_{m_1,i_1} \otimes \cdots \otimes h_{m_d,i_d}$, and

$$h_{m,i}(t) = \begin{cases} 2^{m/2}, & t \in [(i-1)2^{-m}, (i-1/2)2^{-m}), \\ -2^{m/2}, & t \in [(i-1/2)2^{-m}, i2^{-m}), \\ 0, & \text{otherwise.} \end{cases}$$

Case $d > 2$. Note that

(4.3)
$$f_i - f_j = \frac{1}{\sqrt{|\mathcal{M}_n|}} \sum_{M \in \mathcal{M}_n} \sum_{I \in \mathcal{I}_M} (\eta_{M,I}^{(i)} - \eta_{M,I}^{(j)})\Phi_{M,I},$$

where

$$\Phi_{M,I} = \phi_{m_1,i_1} \otimes \cdots \otimes \phi_{m_d,i_d},$$

and

$$\phi_{m,i}(t) = \begin{cases} 0, & t \notin ((i-1)2^{-m}, i2^{-m}), \\ 2^{-m/2-2}, & t = (i-3/4)2^{-m}, \\ -2^{-m/2-2}, & t = (i-1/4)2^{-m}, \\ \text{linear}, & \text{otherwise.} \end{cases}$$

($\phi_{m,i}$ is the integral of $\psi_{m,i}$.) Because

$$\langle \phi_{m,i}, h_{p,j}\rangle = \begin{cases} 0, & (p,j) \neq (m,i), \\ 2^{-3m/2-3}, & (p,j) = (m,i), \end{cases}$$

we have

(4.4)
$$\langle \Phi_{M',I'}, H_{M,I}\rangle = \begin{cases} 2^{-3n/2-3d}, & (M,I) = (M',I'), \\ 0, & (M,I) \neq (M',I'). \end{cases}$$

Thus,

$$\langle f_i - f_j, H_{M,I}\rangle = \frac{1}{\sqrt{|\mathcal{M}_n|}}(\eta_{M,I}^{(i)} - \eta_{M,I}^{(j)}) \cdot 2^{-3n/2-3d}.$$

Hence,

$$\begin{aligned} \|f_i - f_j\|_2^2 &\geq \sum_{M \in \mathcal{M}_n} \sum_{I \in \mathcal{I}_M} |\langle f_i - f_j, H_{M,I}\rangle|^2 \\ &\geq \sum_{M \in \mathcal{M}_n} \sum_{I \in \mathcal{I}_M} \frac{2^{-3n-6d}}{|\mathcal{M}_n|}|\eta_{M,I}^{(i)} - \eta_{M,I}^{(j)}|^2 \\ &= \frac{2^{-3n-6d}}{|\mathcal{M}_n|} \cdot 4d_h(\eta^{(i)}, \eta^{(j)}) \geq c_3^2 2^{-2n}, \end{aligned}$$

proving (4.2) for the case $d > 2$.

Case $d = 2$. We multiply out the product in $R_n(\cdot, \eta^{(i)})$ and write

$$
\begin{aligned}
f_i \quad &= \quad \left(x_1 x_2 + \sum_{M \in \mathcal{M}_n} \sum_{I \in \mathcal{I}_M} \eta_{M,I}^{(i)} \Phi_{M,I} \right) \\
&\quad + \sum_{p \geq 2} \sum \int_0^{x_1} \int_0^{x_2} F_{k_1, n-k_1}^{(i)} \cdots F_{k_p, n-k_p}^{(i)} \, dt_1 dt_2 \\
&=: \quad \left(x_1 x_2 + \sum_{M \in \mathcal{M}_n} \sum_{I \in \mathcal{I}_M} \eta_{M,I}^{(i)} \Phi_{M,I} \right) + Q_n^{(i)},
\end{aligned}
$$

where the inner summation in $Q_n^{(i)}$ is over all even integers $0 \leq k_1 < k_2 < \cdots < k_p \leq n$. We show that for each $M = (m_1, m_2) \in \mathcal{M}$ and $I \in \mathcal{I}_M$, $\left\langle Q_n^{(i)}, H_{M,I} \right\rangle = 0$. Indeed, we can write Q_n as

$$
\sum_{p \geq 2} \sum (\pm) \int_0^{x_1} [\psi_{k_1, i_1}(t_1) \cdots \psi_{k_p, i_p}(t_1)] dt_1 \cdot \int_0^{x_2} [\psi_{n-k_1, j_1}(t_2) \cdots \psi_{n-k_p, j_p}(t_2)] dt_2,
$$

where the inner summation is over all even integers $0 \leq k_1 < k_2 \leq \cdots < k_p \leq n$ and all $1 \leq i_l \leq 2^{k_l}$, $1 \leq j_l \leq 2^{n-k_l}$, $1 \leq l \leq p$. Because $\psi_{k_1, i_1}, \ldots, \psi_{k_{p-1}, i_{p-1}}$ is constant on the support of ψ_{k_p, i_p}, we have

$$
\int_0^{x_1} [\psi_{k_1, i_1}(t_1) \cdots \psi_{k_p, i_p}(t_1)] dt_1 = \text{const} \cdot \phi_{k_p, i_p}(x_1).
$$

Similarly,

$$
\int_0^{x_2} [\psi_{n-k_1, j_1}(t_2) \cdots \psi_{n-k_p, j_p}(t_2)] dt_2 = \text{const} \cdot \phi_{n-k_1, j_1}(x_2).
$$

Because $k_p + n - k_1 > m_1 + m_2 = n$, we must have $k_p > m_1$ or $n - k_1 > m_2$. If $k_p > m_1$, then for any $1 \leq r_1 \leq 2^{m_1}$, h_{m_1, r_1} is constant on the support of ϕ_{k_p, i_p}, and therefore,

$$
\int_0^1 \phi_{k_p, i_p}(x_1) h_{m_1, r_1}(x_1) dx_1 = \text{const} \cdot \int_0^1 \phi_{k_p, i_p}(x_1) dx_1 = 0.
$$

Similarly, if $n - k_1 > m_2$, then $\int_0^1 \phi_{n-k_1, j_1}(x_2) h_{m_2, r_2}(x_1) dx_1 = 0$. In either case, we have

$$
\int_0^1 \int_0^1 \phi_{k_p, i_p}(x_1) \phi_{n-k_1, j_1}(x_2) \cdot h_{m_1, r_1}(x_1) h_{m_2, r_2}(x_2) dx_1 dx_2 = 0.
$$

Hence, $\left\langle Q_n^{(i)}, H_{M,I} \right\rangle = 0$. Therefore,

$$
\langle f_i - f_j, H_{M,I} \rangle = \sum_{P \in \mathcal{M}_n} \sum_{L \in \mathcal{I}_P} (\eta_{P,L}^{(i)} - \eta_{P,L}^{(j)}) \langle \Phi_{P,L}, H_{M,I} \rangle = (\eta_{M,I}^{(i)} - \eta_{M,I}^{(j)}) 2^{-3n/2 - 3d},
$$

which implies

$$
\begin{aligned}
\|f_i - f_j\|_2^2 \quad &\geq \quad \sum_{M \in \mathcal{M}} \sum_{I \in \mathcal{I}_M} |\eta_{M,I}^{(i)} - \eta_{M,I}^{(j)}|^2 2^{-3n - 6d} \\
&= \quad 2^{-3n - 6d} \cdot 4 d_h(\eta^{(i)}, \eta^{(j)}) \geq c_3^2 n 2^{-2n},
\end{aligned}
$$

and thus (4.2) holds for the case $d = 2$.

Proof of the lower bound in Theorem 1.1. Let

$$\mathcal{G}_d = \left\{ g : g(x_1, x_2, ..., x_d) = \int_0^{x_1} \cdots \int_0^{x_d} r(t_1, ..., t_d) dt_1 \cdots dt_d; \|r\|_1 \le 1 \right\}.$$

Note that we have constructed $\exp(c_1 n^{d-1} 2^n)$ functions $f_i \in \mathcal{G}_d$ with $\|f_i\|_\infty \le 1$ and $\|f_i - f_j\|_2$, which satisfy (4.2) for $i \ne j$. In the case $d > 2$, we choose $\varepsilon = \sqrt{c_3} 2^{-n}$ and obtain

$$(4.5) \qquad \log N(\mathcal{G}_d, \| \cdot \|_2, \varepsilon) \ge c_1 n^{d-1} 2^{-n} \approx c' \varepsilon^{-1} (\log 1/\varepsilon)^{d-1}.$$

In the case $d = 2$, we choose $\varepsilon = \sqrt{c_3 n} 2^{-n}$ and obtain

$$(4.6) \qquad \log N(\mathcal{G}_2, \| \cdot \|_2, \varepsilon) \ge c_1 n 2^{-n} \approx c' \varepsilon^{-1} (\log 1/\varepsilon)^{3/2}.$$

By monotonicity of the covering number, (4.5) and (4.6) hold for all small ε.

Because $\mathcal{G}_d \subset \mathcal{F}_d - \mathcal{F}_d$, we have $N(\mathcal{G}_d, \| \cdot \|_2, \varepsilon) \le [N(\mathcal{F}_d, \| \cdot \|_2, \varepsilon/2)]^2$, and the lower bound in Theorem 1.1 follows.

Remark 4.1. An interesting problem is to find the rate of $\log N(\mathcal{F}_d, \| \cdot \|_p, \varepsilon)$ under the L_p-norm, $2 < p < \infty$. Note also that the set $\mathcal{F}_d$ is not compact under the sup-norm.

References

1. Artstein, S., Milman, V., and Szarek, S. (2004). Duality of metric entropy, *Ann. of Math.* **159**, 1313–1328. MR2113023 (2005h:47037)
2. Artstein, S., Milman, V., Szarek, S. J., and Tomczak-Jaegermann, N. (2004). On Convexified Packing and Duality, *Geom. and Funct. Anal.* **14**, 1134-1141. MR2105957 (2005h:47038)
3. Carl, B. and Stephani, I. (1990). *Entropy, Compactness and Approximation of Operators*, Cambridge University Press, Cambridge. MR1098497 (92e:47002)
4. Dunker, T., Kuhn, T., Lifshits, M. A., and Linde, W. (1999). Metric entropy of integration operators and small ball probabilities for the Brownian sheet, *J. Approx. Theory* **101**, 63–77. MR1724026 (2001d:60032)
5. EDMUNDS, D. E. and TRIEBEL, H. (1996). *Function Spaces, Entropy Numbers and Differential Operators*, Cambridge Univ. Press., Cambridge. MR1410258 (97h:46045)
6. Gao, F. (2004). Entropy of absolute convex hulls in Hilbert spaces, *Bull. London Math. Soc.* **36** (2004), 460 - 468. MR2069008 (2005e:41071)
7. Gao, F. and Li, W.V. (2007). Small ball probabilities for Slepian Gaussian fields, *Trans. Amer. Math. Soc.* **359**, 1339–1350. MR2262853
8. Kolmogorov, A.N. and Tihomirov, V.M. (1961). ε-entropy and ε-capacity of sets in function spaces, *Uspehi Mat. Nauk.* **14** (1959), 3-86. English transl. in *Amer. Math. Soc. Transl.* **17**, 277-364. MR0112032 (22:2890)
9. Kuelbs, J. and Li, W.V. (1993). Metric entropy and the small ball problem for Gaussian measures, *J. Funct. Anal.* **116**, 133-157. MR1237989 (94j:60078)
10. Li, W.V. and Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures, *Ann. Probab.* **27**, 1556-1578. MR1733160 (2001c:60059)
11. Li, W.V. and Shao, Q.-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications, Stochastic processes: theory and methods, Handbook of Statist., Vol. 19. 533–597. MR1861734
12. Lorentz, G. (1966). Metric entropy and approximation, *Bull. Amer. Math. Soc.* **72**, 903–937. MR0203320 (34:3173)
13. Talagrand, M. (1994). The small ball problem for the Brownian sheet, *Ann. Probab.* **22**, 1331–1354. MR1303647 (95k:60049)
14. van der Vaart, A. and Wellner, J. *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York, 1996. MR1385671 (97g:60035)

RON BLEI, FUCHANG GAO, AND WENBO V. LI

Department of Mathematics, University of Connecticut, Storrs, Connecticut 06268
  E-mail address: blei@math.uconn.edu

Department of Mathematics, University of Idaho, Moscow, Idaho 83844
  E-mail address: fuchang@uidaho.edu

Department of Mathematical Sciences, University of Delaware, Newark, Delaware
19716
  E-mail address: wli@math.udel.edu