Covering Numbers for Convex Functions

Adityanand Guntuboyina and Bodhisattva Sen

Abstract—In this paper, we study the covering numbers of the space of convex and uniformly bounded functions in multidimension. We find optimal upper and lower bounds for the ϵ-covering number of $\mathcal{C}([a,b]^d, B)$, in the L_p-metric, $1 \leq p < \infty$, in terms of the relevant constants, where $d \geq 1$, $a < b \in \mathbb{R}$, $B > 0$, and $\mathcal{C}([a,b]^d, B)$ denotes the set of all convex functions on $[a,b]^d$ that are uniformly bounded by B. We summarize previously known results on covering numbers for convex functions and also provide alternate proofs of some known results. Our results have direct implications in the study of rates of convergence of empirical minimization procedures as well as optimal convergence rates in the numerous convexity constrained function estimation problems.

Index Terms—Convexity constrained function estimation, empirical risk minimization, Hausdorff distance, Kolmogorov entropy, L_p-metric, metric entropy, packing numbers.

I. INTRODUCTION

EVER since the work of [1], covering numbers (and their logarithms, known as metric entropy numbers) have been studied extensively in a variety of disciplines. For a subset $\mathcal{F}$ of a metric space $(\mathcal{X}, \rho)$, the ϵ-*covering number* $M(\mathcal{F}, \epsilon; \rho)$ is defined as the smallest number of balls of radius ϵ whose union contains $\mathcal{F}$. Covering numbers capture the *size* of the underlying metric space and play a central role in a number of areas in information theory and statistics, including nonparametric function estimation, density estimation, empirical processes, and machine learning.

In this paper, we study the covering numbers of the space of convex and uniformly bounded functions in multidimension. Specifically, we find optimal upper and lower bounds for the ϵ-covering number $M(\mathcal{C}([a,b]^d, B), \epsilon; L_p)$, in the L_p-metric, $1 \leq p < \infty$, in terms of the relevant constants, where $d \geq 1$, $a, b \in \mathbb{R}$, $B > 0$, and $\mathcal{C}([a,b]^d, B)$ denotes the set of all convex functions on $[a,b]^d$ that are uniformly bounded by B. We also summarize previously known results on covering numbers for convex functions. The special case of the problem when $d = 1$ has been recently established by Dryanov in [2, Th. 3.1]. Prior to [2], the only other result on the covering numbers of convex functions is due to Bronshtein in [3] who considered convex functions that are uniformly bounded and uniformly Lipschitz with a *known* Lipschitz constant under the L_∞ metric.

A. Guntuboyina is with the Department of Statistics, University of California, Berkeley, CA 94720 USA (e-mail: aditya@stat.berkeley.edu).

B. Sen is with the Department of Statistics, Columbia University, New York, NY 10027 USA (e-mail: bodhi@stat.columbia.edu).

As will be clear from our proof techniques, results on the covering numbers of convex sets are quite relevant to this paper for which the two main references are [3] and [4, Sec. 4]. The main result in [4, Sec. 4] is actually weaker compared to [3, Th. 5]. However, there are a few minor errors in the proof of [3, Th. 5] which are corrected in the exposition of [5, Ch. 8]. Our results are analogous to the results on covering numbers of classes of smooth functions for which two main references are [1] and [6].

In recent years, there has been an upsurge of interest in nonparametric function estimation under convexity-based constraints, especially in multidimension. In general function estimation, it is well known (see, e.g., [7]–[10]) that the covering numbers of the underlying function space can be used to characterize optimal rates of convergence. They are also useful for studying the rates of convergence of empirical minimization procedures (see, e.g., [11] and [12]). Our results have direct implications in this regard in the context of understanding the rates of convergence of the numerous convexity constrained function estimators, e.g., the nonparametric least squares estimator of a convex regression function studied in [13] and [14]; and the maximum likelihood estimator of a log-concave density in multidimension studied in [15]–[17]. Also, similar problems that crucially use convexity/concavity constraints to estimate sets have also received recent attention in the statistical and machine learning literature, see, e.g., [18], [19], and our results can be applied in such settings.

This paper is organized as follows. In Section II, we set up notation and provide motivation for our main results, which are proved in Section III. In Section IV, we draw some connections to previous results on covering numbers for convex functions and prove a related auxiliary result along with some inequalities of possible independent interest.

II. MOTIVATION

The first result on covering numbers for convex functions was proved by Bronshtein in [3], who considered convex functions defined on a cube in $\mathbb{R}^d$ that are uniformly bounded and uniformly Lipschitz. Specifically, let $\mathcal{C}([a,b]^d, B, \Gamma)$ denote the class of real-valued convex functions defined on $[a,b]^d$ that are uniformly bounded in absolute value by B and uniformly Lipschitz with constant Γ. In Theorem 6 of [3], Bronshtein proved that for ϵ sufficiently small, the logarithm of $M(\mathcal{C}([a,b]^d, B, \Gamma), \epsilon; L_\infty)$ can be bounded from above and below by a positive constant (not depending on ϵ) multiple of $\epsilon^{-d/2}$. Note that the L_∞ distance between two functions f and g on $[a,b]^d$ is defined as $\|f-g\|_\infty := \sup_{x \in [a,b]^d} |f(x) - g(x)|$.

Bronshtein worked with the class $\mathcal{C}([a,b]^d, B, \Gamma)$ where the functions are uniformly Lipschitz with constant Γ. However, in convexity-based function estimation problems, one usually does not have a known uniform Lipschitz bound on the unknown

function class. This leads to difficulties in the analysis of empirical minimization procedures via Bronshtein's result. To the best of our knowledge, there does not exist any other result on the covering numbers of convex functions that deals with all $d \geq 1$ and does not require the Lipschitz constraint.

In the absence of the uniformly Lipschitz constraint (i.e., if one works with the class $\mathcal{C}([a, b]^d, B)$ instead of $\mathcal{C}([a, b]^d, B, \Gamma)$), the covering numbers under the L_∞ metric are infinite. In other words, the space $\mathcal{C}([a, b]^d, B)$ is not totally bounded under the L_∞ metric. This can be seen, for example, by noting that the functions

$$f_j(t) := \max\left(0, 1 - 2^j t\right), \qquad \text{for } t \in [0, 1]$$

are in $\mathcal{C}([0, 1], 1)$ for all $j \geq 1$, and satisfy

$$\|f_j - f_k\|_\infty \geq |f_j(2^{-k}) - f_k(2^{-k})| = 1 - 2^{j-k} \geq 1/2$$

for all $j < k$.

This motivated us to study the covering numbers of the class $\mathcal{C}([a, b]^d, B)$ under a different metric, namely the L_p-metric for $1 \leq p < \infty$. We recall that under the L_p-metric, $1 \leq p < \infty$, the distance between two functions f and g on $[a, b]^d$ is defined as

$$\|f - g\|_p := \left(\int_{x \in [a, b]^d} |f(x) - g(x)|^p dx\right)^{1/p}.$$

Our main result in this paper shows that if one works with the L_p-metric as opposed to L_∞, then the covering numbers of $\mathcal{C}([a, b]^d, B)$ are finite. Moreover, they are bounded from above and below by constant multiples of $\epsilon^{-d/2}$ for sufficiently small ϵ.

III. L_p–Covering Number Bounds for $\mathcal{C}([a, b]^d, B)$

In this section, we prove upper and lower bounds for the ϵ-covering number of $\mathcal{C}([a, b]^d, B)$ under the L_p-metric, $1 \leq p \leq \infty$. Let us start by noting a simple scaling identity that allows us to take $a = 0, b = 1$, and $B = 1$, without loss of generality. For each $f \in \mathcal{C}([a, b]^d, B)$, let us define $\tilde{f}$ on $[0, 1]^d$ by $\tilde{f}(x) := f(a\mathbf{1} + (b - a)x)/B$, where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$. Clearly, $\tilde{f} \in \mathcal{C}([0, 1]^d, 1)$ and, for $1 \leq p < \infty$

$$B^p \int_{x \in [0, 1]^d} \left|\tilde{f}(x) - g(x)\right|^p dx$$
$$= (b - a)^{-d} \int_{y \in [a, b]^d} \left|f(y) - Bg\left(\frac{y - a\mathbf{1}}{b - a}\right)\right|^p dy$$

for $g \in \mathcal{C}([0, 1]^d, 1)$. It follows that covering f to within ϵ in the L_p-metric on $[a, b]^d$ is equivalent to covering $\tilde{f}$ to within $(b - a)^{-d/p} \epsilon / B$ in the L_p-metric on $[0, 1]^d$. Therefore, for $1 \leq p < \infty$

$$M(\mathcal{C}([a, b]^d, B), \epsilon; L_p) = M(\mathcal{C}([0, 1]^d, 1), \epsilon'; L_p) \quad (1)$$

where $\epsilon' := (b - a)^{-d/p} \epsilon / B$.

A. Upper Bound for $M(\mathcal{C}([a, b]^d, B), \epsilon; L_p)$

Theorem 3.1: Fix $1 \leq p < \infty$. There exist positive constants c and ϵ_0, depending only on the dimension d and p, such that, for every $B > 0$ and $b > a$, we have

$$\log M\left(\mathcal{C}([a, b]^d, B), \epsilon; L_p\right) \leq c\left(\frac{\epsilon}{B(b - a)^{d/p}}\right)^{-d/2}$$

for every $\epsilon \leq \epsilon_0 B(b - a)^{d/p}$.

The main ingredient in our proof of the aforementioned theorem is an extension of Bronshtein's theorem to uniformly bounded convex functions having different Lipschitz constraints in different directions. Specifically, for $B \in (0, \infty)$, $\Gamma_i \in (0, \infty]$, and $a_i < b_i$ for $i = 1, \ldots, d$, let $\mathcal{C}\left(\prod_{i=1}^d [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d\right)$ denote the set of all real-valued convex functions f on the rectangle $[a_1, b_1] \times \cdots \times [a_d, b_d]$ that are uniformly bounded by B and satisfy

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d)$$
$$- f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d)| \leq \Gamma_i |x_i - y_i| \quad (2)$$

for every $i = 1, \ldots, d$; $x_i, y_i \in [a_i, b_i]$ and $x_j \in [a_j, b_j]$ for $j \neq i$. In other words, the function $x \mapsto f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d)$ is Lipschitz on $[a_i, b_i]$ with constant Γ_i for all $x_j \in [a_j, b_j], j \neq i$.

Clearly, the class $\mathcal{C}([a, b]^d, B, \Gamma)$ that Bronshtein studied is contained in $\mathcal{C}([a, b]^d; B; \Gamma, \ldots, \Gamma)$. Also, it is easy to check that every function f in $\mathcal{C}\left(\prod_i [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d\right)$ is Lipschitz with respect to the Euclidean norm on $\prod_i [a_i, b_i]$ with Lipschitz constant $\sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2}$.

Note that for $\Gamma_i = \infty$, the inequality (2) is satisfied by every function f. As a result, we have the equality $\mathcal{C}([a, b]^d, B) = \mathcal{C}([a, b]^d; B; \infty, \ldots, \infty)$. The following result gives an upper bound for the ϵ-covering number of $\mathcal{C}(\prod_i [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d)$ and is the main ingredient in the proof of Theorem 3.1. Its proof is similar to Bronshtein's proof [3, Proof of Theorem 6] of his upper bound on $\mathcal{C}([a, b]^d, B, \Gamma)$ and is included in Section IV.

Theorem 3.2: There exist positive constants c and ϵ_0, depending only on the dimension d, such that for every positive $B, \Gamma_1, \ldots, \Gamma_d$ and rectangle $[a_1, b_1] \times \cdots \times [a_d, b_d]$, we have

$$\log M\left(\mathcal{C}\left(\prod_{i=1}^d [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d\right), \epsilon; L_\infty\right)$$
$$\leq c\left(\frac{B + \sum_{i=1}^d \Gamma_i(b_i - a_i)}{\epsilon}\right)^{d/2} \quad (3)$$

for all $0 < \epsilon \leq \epsilon_0\{B + \sum_{i=1}^d \Gamma_i(b_i - a_i)\}$.

Remark 3.1: Note that the right-hand side of (3) equals ∞ unless $\Gamma_i < \infty$ for all $i = 1, \ldots, d$. Thus, Theorem 3.2 is only meaningful when $\Gamma_i < \infty$ for all $i = 1, \ldots, d$.

Remark 3.2: Because $\mathcal{C}([a, b]^d, B, \Gamma)$ is contained in $\mathcal{C}([a, b]^d; B; \Gamma_1, \ldots, \Gamma_d)$, Theorem 3.2 includes Bronshtein's upper bound on $\mathcal{C}([a, b]^d, B, \Gamma)$ as a special case. Moreover, it gives explicit dependence of the upper bound on the constants

a, b, B, and Γ. Bronshtein did not state the dependence on these constants.

We are now ready to prove Theorem 3.1 using Theorem 3.2. Here is the intuition behind the proof. The class $\mathcal{C}([a, b]^d, B)$ can be thought of as an expansion of the class $\mathcal{C}([a, b]^d; B; \Gamma_1, \ldots, \Gamma_d)$ formed by the removal of the d Lipschitz constraints $\Gamma_1, \ldots, \Gamma_d$ (or equivalently, by setting $\Gamma_1 = \cdots = \Gamma_d = \infty$). Instead of removing all these d Lipschitz constraints at the same time, we remove them sequentially one at a time. This is formally accomplished by induction on the number of indices i for which $\Gamma_i = \infty$. Each step of the induction argument focuses on the removal of one finite Γ_i and is thus like solving the 1-D problem. We consequently use Dryanov's ideas from [2, Th. 3.1] to solve this quasi 1-D problem which allows us to complete the induction step.

Proof of Theorem 3.1: The scaling identity (1) lets us take $a = 0, b = 1$, and $B = 1$.

We shall prove that there exist positive constants c and ϵ_0, depending only on d and p, such that for every $\Gamma_i \in (0, \infty]$, we have

$$\log M \left(\mathcal{C}\left([0, 1]^d; 1; \Gamma_1, \ldots, \Gamma_d\right); \epsilon; L_p \right)$$
$$\leq c \left(\frac{2 + \sum_{i=1}^d \Gamma_i \{\Gamma_i < \infty\}}{\epsilon} \right)^{d/2} \quad (4)$$

for $0 < \epsilon \leq \epsilon_0$. Note that this proves the theorem because we can set $\Gamma_i = \infty$ for all $i = 1, \ldots, d$. Our proof will involve induction on l: the number of indices i for which $\Gamma_i = \infty$.

For $l = 0$, i.e., when $\Gamma_i < \infty$ for all $i = 1, \ldots, d$, (4) is a direct consequence of Theorem 3.2. In fact, in this case, (4) also holds for $p = \infty$. Suppose now that (4) holds for all $l < k$ for some $k \in \{1, \ldots, d\}$. We shall then verify it for $l = k$. Fix $\Gamma_i \in (0, \infty]$ such that exactly k of them equal infinity. Without loss of generality, we assume that $\Gamma_1 = \cdots = \Gamma_k = \infty$ and $\Gamma_i < \infty$ for $i > k$. For every sufficiently small $\epsilon > 0$, we shall exhibit an ϵ-cover of $\mathcal{C}([0, 1]^d; 1; \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d)$ in the L_p-metric whose cardinality has logarithm bounded from above by a constant multiple of $(\sum_{i>k} \Gamma_i + 2)^{d/2} \epsilon^{-d/2}$. Note that for $k = d$, the term $\sum_{i>k} \Gamma_i$ equals zero. For convenience, let us denote the class $\mathcal{C}([0, 1]^d; 1; \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d)$ by $\mathcal{G}$ in the rest of this proof.

Let

$$u := \exp\left(-2(p+1)^2(p+2)\log 2\right) \quad \text{and} \quad v := 1 - u. \quad (5)$$

Fix $\eta > 0$ and choose an integer A and $\delta_1, \ldots, \delta_{A+1}$ such that

$$\eta^p = \delta_1 < \cdots < \delta_A < u \leq \delta_{A+1}.$$

For every two functions f and g on $[0, 1]^d$, we can obviously decompose the integral $\int |f - g|^p$ as

$$\int_{[0,1]^d} |f - g|^p = \int_{[0,u] \times [0,1]^{d-1}} |f - g|^p$$
$$+ \int_{[u,v] \times [0,1]^{d-1}} |f - g|^p + \int_{[v,1] \times [0,1]^{d-1}} |f - g|^p.$$

Also

$$\int_{[0,u] \times [0,1]^{d-1}} |f - g|^p \leq \int_{[0,\delta_1] \times [0,1]^{d-1}} |f - g|^p$$
$$+ \sum_{m=1}^A \int_{[\delta_m, \delta_{m+1}] \times [0,1]^{d-1}} |f - g|^p.$$

For a fixed $m = 1, \ldots, A$, consider the problem of covering the functions in $\mathcal{G}$ on the rectangular strip $[\delta_m, \delta_{m+1}] \times [0, 1]^{d-1}$. Clearly

$$\int_{[\delta_m, \delta_{m+1}] \times [0,1]^{d-1}} |f - g|^p = (\delta_{m+1} - \delta_m) \int_{[0,1]^d} |\tilde{f} - \tilde{g}|^p \quad (6)$$

where for $x = (x_1, \ldots, x_d) \in [0, 1]^d$

$$\tilde{f}(x) := f(\delta_m + (\delta_{m+1} - \delta_m)x_1, x_2, \ldots, x_d),$$
$$\text{and} \quad \tilde{g}(x) := g(\delta_m + (\delta_{m+1} - \delta_m)x_1, x_2, \ldots, x_d).$$

By convexity, the restriction of every function f in $\mathcal{G}$ to $[\delta_m, \delta_{m+1}] \times [0, 1]^{d-1}$ belongs to the class

$$\mathcal{C}([\delta_m, \delta_{m+1}] \times [0, 1]^{d-1}; 1; 2/\delta_m, \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d).$$

Consequently, the corresponding function $\tilde{f}$ belongs to

$$\mathcal{C}([0, 1]^d; 1; 2(\delta_{m+1} - \delta_m)/\delta_m, \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d).$$

Because $2(\delta_{m+1} - \delta_m)/\delta_m < \infty$, we can use the induction hypothesis to assert the existence of positive constants ϵ_0 and c, depending only on d and p, such that for every positive real number $\alpha_m \leq \epsilon_0$, there exists an α_m-cover of $\mathcal{C}([0, 1]^d; 1; 2(\delta_{m+1} - \delta_m)/\delta_m, \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d))$ in the L_p-metric on $[0, 1]^d$ of size smaller than

$$\exp\left(c\alpha_m^{-d/2} \left(2 + \frac{2(\delta_{m+1} - \delta_m)}{\delta_m} + \sum_{i>k} \Gamma_i \right)^{d/2} \right)$$
$$\leq \exp\left(c\left(2 + \sum_{i>k} \Gamma_i \right)^{d/2} \left(\frac{\delta_{m+1}}{\delta_m \alpha_m} \right)^{d/2} \right).$$

By covering the functions in $\mathcal{G}$ by the constant function 0 on $[0, \delta_1] \times [0, 1]^{d-1}$ and up to α_m in the L_p-metric on $[\delta_m, \delta_{m+1}] \times [0, 1]^{d-1}$ for $m = 1, \ldots, A$, we obtain a cover of the restriction of the functions in $\mathcal{G}$ to the set $[0, u] \times [0, 1]^{d-1}$ in L_p-metric having coverage $S_1^{1/p}$ and cardinality bounded from above by $\exp(S_2)$ where

$$S_1 := \delta_1 + \sum_{m=1}^A \alpha_m^p (\delta_{m+1} - \delta_m) \quad \text{and}$$

$$S_2 := c\left(\sum_{i>k} \Gamma_i + 2 \right)^{d/2} \sum_{m=1}^A \left(\frac{\delta_{m+1}}{\delta_m \alpha_m} \right)^{d/2}. \quad (7)$$

Suppose now that

$$\delta_m := \exp\left(p\left(\frac{p+1}{p+2}\right)^{m-1}\log\eta\right) \quad \text{and}$$

$$\alpha_m := \eta\exp\left(-p\frac{(p+1)^{m-2}}{(p+2)^{m-1}}\log\eta\right)$$

for $m = 1, \ldots, A+1$, where A is the largest integer such that

$$\exp\left(p\left(\frac{p+1}{p+2}\right)^{A-1}\log\eta\right) < u.$$

Then

$$S_1 = \delta_1 + \sum_{m=1}^{A}\alpha_m^p\left(\delta_{m+1} - \delta_m\right)$$

$$\leq \delta_1 + \sum_{m=1}^{A}\alpha_m^p\delta_{m+1} = \eta^p\left(1 + \sum_{m=1}^{A}\zeta_m^2\right)$$

and

$$S_2 = c\left(\frac{\sum_{i>k}\Gamma_i + 2}{\eta}\right)^{d/2}\sum_{m=1}^{A}\zeta_m^d$$

where

$$\zeta_m := \sqrt{\frac{\eta\delta_{m+1}}{\delta_m\alpha_m}} = \exp\left(\frac{p}{2(p+1)^2}\frac{(p+1)^m}{(p+2)^m}\log\eta\right).$$

Note that if $\eta \leq 1$, then $\log\eta \leq 0$ which implies $\zeta_m \leq 1$. Also, for $m = 2, \ldots, A$, we have

$$\frac{\zeta_m}{\zeta_{m-1}} = \exp\left(\frac{-p\log\eta}{2(p+1)^2(p+2)}\left(\frac{p+1}{p+2}\right)^{m-1}\right)$$

$$\geq \exp\left(\frac{-p\log\eta}{2(p+1)^2(p+2)}\left(\frac{p+1}{p+2}\right)^{A-1}\right)$$

$$= \exp\left(\frac{-\log\delta_A}{2(p+1)^2(p+2)}\right)$$

$$> \exp\left(\frac{-\log u}{2(p+1)^2(p+2)}\right) = 2$$

where we have used $\delta_A < u$ and the fact that u has the expression (5). Therefore, $\zeta_m \geq 2\zeta_{m-1}$ which can be rewritten as

$$\zeta_m^r \leq \frac{2^r}{2^r - 1}\left(\zeta_m^r - \zeta_{m-1}^r\right) \qquad \text{for every } r \geq 1.$$

Thus

$$\sum_{m=1}^{A}\zeta_m^r \leq \zeta_1^r + \frac{2^r}{2^r - 1}\sum_{m=2}^{A}\left(\zeta_m^r - \zeta_{m-1}^r\right)$$

$$= \frac{1}{2^r - 1}\left(2^r\zeta_A^r - \zeta_1^r\right) \leq \frac{2^r}{2^r - 1}.$$

Using this for $r = 2$ and $r = d$, we deduce that

$$S_1 \leq \frac{7}{3}\eta^p \quad \text{and} \quad S_2 \leq \frac{2^d c}{2^d - 1}\left(\frac{\sum_{i>k}\Gamma_i + 2}{\eta}\right)^{d/2}.$$

An exactly similar analysis can be done now to cover the restrictions of the functions in $\mathcal{G}$ to the set $[v, 1] \times [0, 1]^{d-1}$ having the same coverage $S_1^{1/p}$ and same cardinality bounded by $\exp(S_2)$. For $[u, v] \times [0, 1]^{d-1}$, we note, by convexity, that the restrictions of functions in $\mathcal{G}$ to the set $[u, v] \times [0, 1]^{d-1}$ belong to $\mathcal{C}([u, v] \times [0, 1]^{d-1}; 1; 2/u, \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d)$. By the induction hypothesis, there exist constants c and ϵ_0, depending only on d and p, such that for all $\eta \leq \epsilon_0$, one can get a ϵ-cover of $\mathcal{C}([u, v] \times [0, 1]^{d-1}; 1; 2/u, \infty, \ldots, \infty, \Gamma_{k+1}, \ldots, \Gamma_d)$ in the L_p-metric having cardinality smaller than

$$\exp\left(c\eta^{-d/2}\left(2 + \frac{2}{u} + \sum_{i>k}\Gamma_i\right)^{d/2}\right)$$

$$\leq \exp\left(c\left(\frac{2}{u}\right)^{d/2}\left(\frac{\sum_{i>k}\Gamma_i + 2}{\eta}\right)^{d/2}\right).$$

Observe that u only depends on p. By combining the covers of the restrictions of functions in $\mathcal{G}$ to these three strips $[0, u] \times [0, 1]^{d-1}$, $[u, v] \times [0, 1]^{d-1}$, and $[v, 1] \times [0, 1]^{d-1}$, we obtain, for $\eta \leq \epsilon_0$, a cover of $\mathcal{G}$ in the L_p-metric having coverage at most

$$\left(\frac{7}{3}\eta^p + \frac{7}{3}\eta^p + \eta^p\right)^{1/p} = \left(\frac{17}{3}\right)^{1/p}\eta$$

and cardinality at most

$$\exp\left(c\left(\frac{2^{d+1}}{2^d - 1} + \frac{2^{d/2}}{u^{d/2}}\right)\left(\frac{\sum_{i>k}\Gamma_i + 2}{\eta}\right)^{d/2}\right).$$

By relabeling $(17/3)^{1/p}\eta$ as ϵ, we have proved that for $\epsilon \leq (3/17)^{1/p}\epsilon_0$

$$\log M(\mathcal{G}; \epsilon; L_p)$$
$$\leq c\left(\frac{17}{3}\right)^{d/(2p)}\left(\frac{2^{d+1}}{2^d - 1} + \frac{2^{d/2}}{u^{d/2}}\right)\left(\frac{\sum_{i>k}\Gamma_i + 2}{\epsilon}\right)^{d/2}.$$

This proves (4) for all $\Gamma_1, \ldots, \Gamma_d$ such that exactly k of them equal ∞. The proof is complete by induction. ∎

Remark 3.3: The argument used in the aforementioned induction step involved splitting the interval $[0, 1]$ into the three intervals $[0, u]$, $[u, v]$, and $[v, 1]$, and then subsequently splitting the interval $[0, u]$ into smaller subintervals. We have borrowed this idea from Dryanov [2, Proof of Theorem 3.1]. We must mention however that Dryanov uses a more elaborate argument to bound sums of the form S_1 and S_2. Our way of controlling S_1 and S_2 is much simpler which shortens the argument considerably.

B. Lower Bound for $M(\mathcal{C}([a,b]^d, B), \epsilon; L_p)$

Theorem 3.3: There exist positive constants c and ϵ_0, depending only on the dimension d, such that for every $p \geq 1$, $B > 0$, and $b > a$, we have

$$\log M\left(\mathcal{C}([a,b]^d, B), \epsilon; L_p\right) \geq c\left(\frac{\epsilon}{B(b-a)^{d/p}}\right)^{-d/2}$$

for $\epsilon \leq \epsilon_0 B(b-a)^{d/p}$.

Proof: As before, by the scaling identity (1), we take $a = 0$, $b = 1$, and $B = 1$. For functions defined on $[0,1]^d$, the L_p-metric, $p > 1$, is larger than L_1. We will thus take $p = 1$ in the rest of this proof. We prove that for ϵ sufficiently small, there exists an ϵ-packing subset of $\mathcal{C}([0,1]^d, 1)$, under the L_1-metric, of cardinality larger than a constant multiple of $\epsilon^{-d/2}$. By a packing subset of $\mathcal{C}([0,1]^d, 1)$, we mean a subset F satisfying $\|f - g\|_1 \geq \epsilon$ whenever $f, g \in F$ with $f \neq g$.

Fix $0 < \eta \leq 4(2 + \sqrt{d-1})^{-2}$ and let $k := k(\eta)$ be the positive integer satisfying

$$k \leq \frac{2\eta^{-1/2}}{2 + \sqrt{d-1}} < k + 1 \leq 2k. \qquad (8)$$

Consider the intervals $I(i) = [u(i), v(i)]$ for $i = 1, \ldots, k$, such that

1) $0 \leq u(1) < v(1) \leq u(2) < v(2) \leq \cdots \leq u(k) < v(k) \leq 1$;
2) $v(i) - u(i) = \sqrt{\eta}$, for $i = 1, \ldots, k$;
3) $u(i+1) - v(i) = \frac{1}{2}\sqrt{\eta(d-1)}$ for $i = 1, \ldots, k-1$.

Let $\mathcal{S}$ denote the set of all d-dimensional cubes of the form $I(i_1) \times \cdots \times I(i_d)$ where $i_1, \ldots, i_d \in \{1, \ldots, k\}$. The cardinality of $\mathcal{S}$, denoted by $|\mathcal{S}|$, is clearly k^d.

For each $S \in \mathcal{S}$ with $S = I(i_1) \times \cdots \times I(i_d)$ where $I(i_j) = [u(i_j), v(i_j)]$, let us define the function $h_S : [0,1]^d \to \mathbb{R}$ as

$$h_S(x) = h_S(x_1, \ldots, x_d)$$

$$:= \frac{1}{d}\sum_{j=1}^d \left[u^2(i_j) + \{v(i_j) + u(i_j)\}\{x_j - u(i_j)\}\right]$$

$$= f_0(x) + \frac{1}{d}\sum_{j=1}^d \{x_j - u(i_j)\}\{v(i_j) - x_j\} \qquad (9)$$

where $f_0(x) := \frac{1}{d}\left(x_1^2 + \cdots + x_d^2\right)$, for $x \in [0,1]^d$. The functions $h_S, S \in \mathcal{S}$ have the following four key properties.

1) h_S is affine and hence convex.
2) For every $x \in [0,1]^d$, we have $h_S(x) \leq h_S(1, \ldots, 1) \leq 1$.
3) For every $x \in S$, we have $h_S(x) \geq f_0(x)$. This is because whenever $x \in S$, we have $u(i_j) \leq x_j \leq v(i_j)$ for each j, which implies $\{x_j - u(i_j)\}\{v(i_j) - x_j\} \geq 0$.
4) Let $S, S' \in \mathcal{S}$ with $S \neq S'$. For every $x \in S'$, we have $h_S(x) \leq f_0(x)$. To see this, let $S' = I(i'_1) \times \cdots \times I(i'_d)$ with $I(i'_j) = [u(i'_j), v(i'_j)]$. Let $x \in S'$ and fix $1 \leq j \leq d$. If $I(i_j) = I(i'_j)$, then $x_j \in I(i_j) = [u(i_j), v(i_j)]$ and hence

$$\{x_j - u(i_j)\}\{v(i_j) - x_j\} \leq \frac{\{v(i_j) - u(i_j)\}^2}{4} = \frac{\eta}{4}.$$

If $I(i_j) \neq I(i'_j)$ and $u(i'_j) < v(i'_j) < u(i_j) < v(i_j)$, then

$$\{x_j - u(i_j)\}\{v(i_j) - x_j\}$$
$$\leq -\{u(i_j) - v(i'_j)\}^2 = -\frac{d-1}{4}\eta.$$

The same above bound holds if $u(i_j) < v(i_j) < u(i'_j) < v(i'_j)$. Because $S \neq S'$, at least one of i_j and i'_j will be different. Consequently

$$h_S(x) = f_0(x) + \sum_j \{x_j - u(i_j)\}\{v(i_j) - x_j\}$$

$$\leq f_0(x) + \sum_{j:i_j = i'_j} \frac{\eta}{4} - \sum_{j:i_j \neq i'_j} (d-1)\frac{\eta}{4} \leq f_0(x).$$

Let $\{0,1\}^\mathcal{S}$ denote the collection of all $\{0,1\}$-valued functions on $\mathcal{S}$. The cardinality of $\{0,1\}^\mathcal{S}$ clearly equals $2^{|\mathcal{S}|}$ (recall that $|\mathcal{S}| = k^d$).

For each $\theta \in \{0,1\}^\mathcal{S}$, let

$$g_\theta(x) := \max\left(\max_{S \in \mathcal{S}: \theta(S) = 1} h_S(x), f_0(x)\right).$$

The first two properties of $h_S, S \in \mathcal{S}$ ensure that $g_\theta \in \mathcal{C}([0,1]^d, 1)$. The last two properties imply that

$$g_\theta(x) = h_S(x)\theta(S) + f_0(x)(1 - \theta(S)) \qquad \text{for } x \in S.$$

We now bound from below the L_1 distance between g_θ and $g_{\theta'}$ for $\theta, \theta' \in \{0,1\}^\mathcal{S}$. Because the interiors of the cubes in $\mathcal{S}$ are all disjoint, we can write

$$\|g_\theta - g_{\theta'}\|_1 \geq \sum_{S \in \mathcal{S}} \int_{x \in S} |g_\theta(x) - g_{\theta'}(x)|\, dx$$

$$= \sum_{S \in \mathcal{S}} \{\theta(S) \neq \theta'(S)\} \int_{x \in S} |h_S(x) - f_0(x)| dx.$$

Note that from (9) and by symmetry, the value of integral

$$\zeta := \int_{x \in S} |h_S(x) - f_0(x)| dx$$

is the same for all $S \in \mathcal{S}$. We have thus shown that

$$\|g_\theta - g_{\theta'}\|_1 \geq \zeta \Upsilon(\theta, \theta') \qquad \text{for all } \theta, \theta' \in \{0,1\}^\mathcal{S}, \quad (10)$$

where $\Upsilon(\theta, \theta') := \sum_{S \in \mathcal{S}} \{\theta(S) \neq \theta'(S)\}$ denotes the Hamming distance.

The quantity ζ can be computed in the following way. Let $S = I(i_1) \times \cdots \times I(i_d)$ where $I(i_j) = [u(i_j), v(i_j)]$. We write

$$\zeta = \int_{u(i_1)}^{v(i_1)} \cdots \int_{u(i_d)}^{v(i_d)} \frac{1}{d}\sum_{j=1}^d \{x_j - u(i_j)\}\{v(i_j) - x_j\} dx_d \ldots dx_1.$$

By the change of variable $y_j = \{x_j - u(i_j)\}/\{v(i_j) - u(i_j)\}$ for $j = 1, \ldots, d$, we get

$$\zeta = \prod_{j=1}^d \{v(i_j) - u(i_j)\} \int_{[0,1]^d} \frac{1}{d}\sum_{j=1}^d \{v(i_j) - u(i_j)\}^2 y_j (1 - y_j) dy.$$

Recalling that $v(i) - u(i) = \sqrt{\eta}$ for all $i \leq 1, \ldots, k$, we get $\zeta = \eta^{d/2}\eta\gamma_d$ where

$$\gamma_d := \int_{[0,1]^d} \frac{1}{d} \sum_{j=1}^{d} y_j(1-y_j)dy.$$

Note that γ_d is a constant that depends on the dimension d alone. Thus, from (10), we deduce

$$\|g_\theta - g_{\theta'}\|_1 \geq \gamma_d \eta^{d/2} \eta \Upsilon(\theta, \theta') \qquad (11)$$

for all $\theta, \theta' \in \{0, 1\}^{\mathcal{S}}$. We now use the Varshamov–Gilbert lemma (see, e.g., [20, Lemma 4.7]) which asserts the existence of a subset W of $\{0, 1\}^{\mathcal{S}}$ with cardinality, $|W| \geq \exp(|\mathcal{S}|/8)$ such that $\Upsilon(\tau, \tau') \geq |\mathcal{S}|/4$ for all $\tau, \tau' \in W$ with $\tau \neq \tau'$. Thus, from (11) and (8), we get that for every $\tau, \tau' \in W$ with $\tau \neq \tau'$

$$\|g_\theta - g_{\theta'}\|_1 \geq \gamma_d \eta^{d/2} \eta \frac{|\mathcal{S}|}{4} = \frac{\gamma_d}{4}\eta^{d/2}\eta k^d \geq c_1 \eta$$

where $c_1 := \frac{\gamma_d}{4}(2 + \sqrt{d-1})^{-d}$. Taking $\epsilon := c_1\eta$, we have obtained for $\epsilon \leq \epsilon_0 := 4c_1(2+\sqrt{d-1})^{-2}$, an ϵ-packing subset of $\mathcal{C}([0,1]^d, 1)$ of size $M := |W|$ where

$$\log M \geq \frac{|\mathcal{S}|}{8} = \frac{k^d}{8} \geq \frac{(2 + \sqrt{d-1})^{-d}}{8}\eta^{-d/2}$$

$$= \frac{c_1^{d/2}}{8(2+\sqrt{d-1})^d}\epsilon^{-d/2} = c\epsilon^{-d/2}$$

where c depends only on the dimension d. This completes the proof. ∎

Remark 3.4: The explicit packing subset constructed in the aforementioned proof consists of functions that can be viewed as perturbations of the quadratic function f_0. Previous lower bounds on the covering numbers of convex functions in [3, Proof of Theorem 6] and [2, Sec. 2] (for $d = 1$) are based on perturbations of a function whose graph is a subset of a sphere, a more complicated convex function than f_0. The perturbations of f_0 in the aforementioned proof can also be used to simplify the lower bound arguments in those papers.

IV. DISTANCES BETWEEN CONVEX FUNCTIONS, AND THEIR EPIGRAPHS

One of the aims of this section is to provide the proof of Theorem 3.2. Our strategy for the proof of Theorem 3.2 is similar to Bronshtein's proof of the upper bound on $M(\mathcal{C}([a,b]^d, B, \Gamma), \epsilon; L_\infty)$. The proof involves the following ingredients.

1) An inequality between the L_∞ distance between two convex functions and the Hausdorff distance between their epigraphs.
2) The result of Bronshtein [3] for the covering numbers of convex sets in the Hausdorff metric.

For a convex function f on $[0, 1]^d$ and $B > 0$, let us define the epigraph $V_f(B)$ of f by

$$V_f(B) := \{(x_1, \ldots, x_d, x_{d+1}) : (x_1, \ldots, x_d) \in [0, 1]^d$$
$$\text{and } f(x_1, \ldots, x_d) \leq x_{d+1} \leq B\}.$$

If $f \in \mathcal{C}([0, 1]^d, B)$, then clearly

$$x_1^2 + \cdots + x_d^2 + x_{d+1}^2 \leq 1 + \cdots + 1 + B^2 = d + B^2$$

for every $(x_1, \ldots, x_{d+1}) \in V_f(B)$. Therefore, for every $f \in \mathcal{C}([0, 1]^d, B)$, its epigraph $V_f(B)$ is contained in the $(d + 1)$-dimensional ball of radius $\sqrt{d + B^2}$ centered at the origin. The following inequality relates the L_∞ distance between two functions in $\mathcal{C}([0, 1]^d; B; \Gamma_1, \ldots, \Gamma_d)$ to the Hausdorff distance between their epigraphs. The Hausdorff distance between two compact, convex sets C and D in Euclidean space is defined by

$$\ell_H(C, D) := \max\left(\sup_{x \in C} \inf_{y \in D} |x - y|, \sup_{x \in D} \inf_{y \in C} |x - y|\right)$$

where $|\cdot|$ denotes Euclidean distance.

Lemma 4.1: For every pair of functions f and g in $\mathcal{C}([0, 1]^d; B; \Gamma_1, \ldots, \Gamma_d)$, we have

$$\|f - g\|_\infty \leq \ell_H(V_f(B), V_g(B))\sqrt{1 + \Gamma_1^2 + \cdots + \Gamma_d^2}.$$

Proof: We can clearly assume that $\Gamma_i < \infty$ for all $i = 1, \ldots, d$. Fix $f, g \in \mathcal{C}([0, 1]^d; B; \Gamma_1, \ldots, \Gamma_d)$ and let $\ell_H(V_f(B), V_g(B)) = \rho$. Fix $x \in [0, 1]^d$ with $f(x) \neq g(x)$. Suppose, without loss of generality, that $f(x) < g(x)$. Now $(x, f(x)) \in V_f(B)$ and because $\ell_H(V_f(B), V_g(B)) = \rho$, there exists $(x', y') \in V_g(B)$ with $|(x, f(x)) - (x', y')| \leq \rho$. Because $f(x) < g(x)$, the point $(x, f(x))$ lies outside $V_g(B)$ and using the convexity of $V_g(B)$ we can take $y' = g(x')$. Therefore

$$0 \leq g(x) - f(x)$$
$$= g(x) - g(x') + g(x') - f(x)$$
$$\leq |x - x'|\sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2} + |g(x') - f(x)|$$
$$\leq \sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2 + 1}\sqrt{|x - x'|^2 + |g(x') - f(x)|^2}$$
$$= \sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2 + 1}\,|(x, f(x)) - (x', y')|$$
$$\leq \rho\sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2 + 1}$$

where the second last inequality follows from the Cauchy–Scwarz (C–S) inequality. Lemma 4.1 now follows because $x \in [0, 1]^d$ is arbitrary in the aforementioned argument. ∎

The proof of Theorem 3.2, given in the following, is based on Lemma 4.1 and the following result on covering numbers of convex sets proved in [3]. For $\Gamma > 0$, let $\mathcal{K}^{d+1}(\Gamma)$ denote the set of all compact, convex subsets of the ball in $\mathbb{R}^{d+1}$ of radius Γ centered at the origin. In Theorem 3 (and Remark 1) of [3], Bronshtein proved that there exist positive constants c and ϵ_0, depending only on d, such that

$$\log M(\mathcal{K}^{d+1}(\Gamma), \epsilon; \ell_H) \leq c\left(\frac{\Gamma}{\epsilon}\right)^{d/2} \qquad \text{for } \epsilon \leq \Gamma\epsilon_0. \quad (12)$$

A more detailed account of Bronshtein's proof of (12) can be found in [5, Sec. 8.4].

Proof of Theorem 3.2: The conclusion of the theorem is clearly only meaningful in the case when $\Gamma_i < \infty$ for all $i = 1, \ldots, d$. We therefore assume this in the rest of this proof.

For every $f \in \mathcal{C}\left(\prod_{i=1}^{d}[a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d\right)$, let us define the function $\hat{f}$ on $[0,1]^d$ by

$$\hat{f}(t_1, \ldots, t_d) := f\left(a_1 + (b_1 - a_1)t_1, \ldots, a_d + (b_d - a_d)t_d\right)$$

for $t_1, t_2, \ldots, t_d \in [0,1]$. Clearly, the function $\hat{f}$ belongs to the class $\mathcal{C}\left([0,1]^d; B; \Gamma_1(b_1 - a_1), \ldots, \Gamma_d(b_d - a_d)\right)$ and covering $\hat{f}$ to within ϵ in the L_∞-metric is equivalent to covering f. Thus

$$M\left(\mathcal{C}(\prod_i [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$= M\left(\mathcal{C}([0,1]^d; B; \Gamma_1(b_1 - a_1), \ldots, \Gamma_d(b_d - a_d)), \epsilon; L_\infty\right). \tag{13}$$

We thus take, without loss of generality, $a_i = 0$ and $b_i = 1$ for all $i = 1, \ldots, d$.

From Lemma 4.1 and the observation that $V_f(B) \in \mathcal{K}^{d+1}(\sqrt{d + B^2})$ for all $f \in \mathcal{C}([0,1]^d, B)$, it follows that

$$M\left(\mathcal{C}([0,1]^d; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$\leq M\left(\mathcal{K}^{d+1}(\sqrt{d + B^2}), \frac{\epsilon}{2\sqrt{1 + \Gamma_1^2 + \cdots + \Gamma_d^2}}; \ell_H\right).$$

Thus, from (12), we deduce the existence of two positive constants c and ϵ_0, depending only on d, such that

$$\log M\left(\mathcal{C}([0,1]^d; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$\leq c\left(\frac{\sqrt{(d + B^2)(1 + \Gamma_1^2 + \cdots + \Gamma_d^2)}}{\epsilon}\right)^{d/2}$$

if $\epsilon \leq \epsilon_0 \sqrt{(d + B^2)(1 + \Gamma_1^2 + \cdots + \Gamma_d^2)}$. By the scaling inequality (13), we obtain

$$\log M\left(\mathcal{C}(\prod_i [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$\leq c\left(\frac{\sqrt{(d + B^2)(1 + \sum_i \Gamma_i^2(b_i - a_i)^2)}}{\epsilon}\right)^{d/2}$$

if $\epsilon \leq \epsilon_0 \sqrt{(d + B^2)(1 + \sum_i \Gamma_i^2(b_i - a_i)^2)}$. By another scaling argument, it follows that

$$M\left(\mathcal{C}(\prod_i [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$= M\left(\mathcal{C}\left(\prod_i [a_i, b_i]; \frac{B}{\Upsilon}; \frac{\Gamma_1}{\Upsilon}, \ldots, \frac{\Gamma_d}{\Upsilon}\right), \frac{\epsilon}{\Upsilon}; L_\infty\right)$$

for every $\Upsilon > 0$ and, as a consequence, we get for every $\Upsilon > 0$

$$\log M\left(\mathcal{C}(\prod_i [a_i, b_i]; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$\leq c\left(\frac{\sqrt{(d\Upsilon^2 + B^2)(1 + \sum_i \Gamma_i^2(b_i - a_i)^2/\Upsilon^2)}}{\epsilon}\right)^{d/2}$$

if $\epsilon \leq \epsilon_0 \sqrt{(d\Upsilon^2 + B^2)(1 + \sum_i \Gamma_i^2(b_i - a_i)^2/\Upsilon^2)}$. Choosing (by differentiation)

$$\Upsilon^4 = \frac{B^2 \sum_i \Gamma_i^2(b_i - a_i)^2}{d}$$

we deduce finally

$$\log M\left(\mathcal{C}([a, b]^d; B; \Gamma_1, \ldots, \Gamma_d), \epsilon; L_\infty\right)$$
$$\leq c\left(\frac{B + \sqrt{d \sum_i \Gamma_i^2(b_i - a_i)^2}}{\epsilon}\right)^{d/2}$$

if $\epsilon \leq \epsilon_0\left(B + \sqrt{d \sum_i \Gamma_i^2(b_i - a_i)^2}\right)$. The proof of the theorem will now be complete by noting that

$$\sqrt{\sum_i \Gamma_i^2(b_i - a_i)^2} \leq \sum_i \Gamma_i(b_i - a_i) \leq \sqrt{d \sum_i \Gamma_i^2(b_i - a_i)^2}.$$

The terms involving d can be absorbed in the constants c and ϵ_0. ∎

One might wonder if a version of Lemma 4.2 can be proved for the L_p-metric instead of the L_∞-metric, and without any Lipschitz constraints. Such an inequality would, in particular, yield an alternative simpler proof of Theorem 3.1. It turns out that one can prove such a bound for the L_1-metric but not for L_p for any $p > 1$. The inequality for L_1 is presented next. This inequality could possibly be of independent interest. The reason why such an inequality cannot be proved for L_p, $p > 1$, is explained in Remark 4.1.

Lemma 4.2: For every pair of functions f and g in $\mathcal{C}([0,1]^d, 1)$, we have

$$\|f - g\|_1 \leq (1 + 20d)\ell_H(V_f(1), V_g(1)). \tag{14}$$

Proof: For $f \in \mathcal{C}([0,1]^d, 1)$ and $x \in (0,1)^d$, let $m_f(x)$ denote any subgradient of the convex function f at x. Let $\ell_H(V_f(1), V_g(1)) = \rho > 0$. Our first step is to observe that

$$|f(x) - g(x)| \leq \rho\left(1 + |m_f(x)| + |m_g(x)|\right) \tag{15}$$

for every $x \in (0,1)^d$, where $|m_f(x)|$ denotes the Euclidean norm of the subgradient vector $m_f(x) \in \mathbb{R}^d$. To see this, fix $x \in (0,1)^d$ with $f(x) \neq g(x)$. We assume, without loss of generality, that $f(x) < g(x)$. Clearly, $(x, f(x)) \in V_f(1)$ and because $\ell_H(V_f(1), V_g(1)) = \rho$, there exists $(x', y') \in V_g(1)$ with $|(x, f(x)) - (x', y')| \leq \rho$. Since $f(x) < g(x)$, the point

$(x, f(x))$ lies outside the convex set $V_g(1)$ and we can thus take $y' = g(x')$. By the definition of the subgradient, we have

$$g(x') \geq g(x) + \langle m_g(x), x' - x \rangle.$$

Therefore

$$
\begin{aligned}
0 \leq g(x) - f(x) = g(x) - g(x') + g(x') - f(x) \\
\leq \langle m_g(x), x - x' \rangle + |g(x') - f(x)| \\
\leq |m_g(x)||x - x'| + |g(x') - f(x)| \\
\leq \sqrt{|m_g(x)|^2 + 1}\, |(x, f(x)) - (x', y')| \\
\leq \rho\sqrt{|m_g(x)|^2 + 1} \leq \rho(1 + |m_g(x)|).
\end{aligned}
$$

Note that the C–S inequality has been used twice in the aforementioned chain of inequalities. We have thus shown that $g(x) - f(x) \leq \rho(1 + |m_g(x)|)$ in the case when $f(x) < g(x)$. One would have a similar inequality in the case when $f(x) > g(x)$. Combining these two, we obtain (15).

As a consequence of (15), we get

$$
\begin{aligned}
\|f - g\|_1 \\
= \int_{[0,1]^d \setminus [\rho, 1-\rho]^d} |f - g| + \int_{[\rho, 1-\rho]^d} |f - g| \\
\leq 2\left(1 - (1-2\rho)^d\right) + \rho\left(1 + \int_{[\rho, 1-\rho]^d} |m_f(x)| dx \right. \\
\left. + \int_{[\rho, 1-\rho]^d} |m_g(x)| dx\right) \\
\leq \rho\left(1 + 4d + \int_{[\rho, 1-\rho]^d} \{|m_f(x)| + |m_g(x)|\} dx\right)
\end{aligned}
$$

where we have used the inequality $(1 - 2\rho)^d \geq 1 - 2d\rho$.

To complete the proof of (14), we show that $\int_{[\rho, 1-\rho]^d} |m_f(x)| dx \leq 8d$ for every $f \in \mathcal{C}([0,1]^d, 1)$. We write $m_f(x) = (m_f(x)(1), \ldots, m_f(x)(d)) \in \mathbb{R}^d$ and use the definition of the subgradient to note that for every $x \in [\rho, 1-\rho]^d$ and $1 \leq i \leq d$

$$f(x + t e_i) - f(x) \geq t\, m_f(x)(i) \qquad (16)$$

for $t > 0$ sufficiently small, where e_i is the unit vector in the ith coordinate direction i.e., $e_i(j) := 1$ if $i = j$ and 0 otherwise. Dividing both sides by t and letting $t \downarrow 0$, we would get $m_f(x)(i) \leq f'(x; e_i)$ (we use $f'(x; v)$ to denote the directional derivative of f in the direction v; directional derivatives exist as f is convex). Using (16) for $t < 0$, we get $m_f(x)(i) \geq -f'(x; -e_i)$. Combining these two inequalities, we get

$$|m_f(x)(i)| \leq |f'(x; e_i)| + |f'(x; -e_i)| \qquad \text{for } i = 1, \ldots, d.$$

As a result

$$
\begin{aligned}
\int_{[\rho, 1-\rho]^d} |m_f(x)| dx \\
\leq \sum_{i=1}^d \int_{[\rho, 1-\rho]^d} |m_f(x)(i)| dx \\
\leq \sum_{i=1}^d \left(\int_{[\rho, 1-\rho]^d} |f'(x; e_i)| dx + \int_{[\rho, 1-\rho]^d} |f'(x; -e_i)| dx\right).
\end{aligned}
$$

We now show that for each i, both the integrals $\int_{[\rho, 1-\rho]^d} |f'(x; e_i)|$ and $\int_{[\rho, 1-\rho]^d} |f'(x; -e_i)|$ are bounded from above by 4. Assume, without loss of generality, that $i = 1$ and notice

$$
\int_{[\rho, 1-\rho]^d} |f'(x; e_1)| dx
$$
$$
= \int_{u \in [\rho, 1-\rho]^{d-1}} \left(\int_\rho^{1-\rho} |f'((x_1, u); e_1)| dx_1\right) du. \quad (17)
$$

We fix $u = (x_2, \ldots, x_d) \in [\rho, 1-\rho]^{d-1}$ and focus on the inner integral. Let $v(z) := f(z, x_2, \ldots, x_d)$ for $z \in [0,1]$. Clearly, v is a convex function on $[0,1]$ and its right derivative, $v_r'(x_1)$ at the point $z = x_1 \in (0,1)$ equals $f'(x; e_1)$ where $x = (x_1, \ldots, x_d)$. The inner integral thus equals $\int_\rho^{1-\rho} |v_r'(z)| dz$. Because of the convexity of v, its right derivative $v_r'(z)$ is nondecreasing and satisfies

$$v(y_2) - v(y_1) = \int_{y_1}^{y_2} v_r'(z) dz \qquad \text{for } 0 < y_1 < y_2 < 1.$$

Consequently

$$
\int_\rho^{1-\rho} |v_r'(z)| dz
$$
$$
\leq \sup_{\rho \leq c \leq 1-\rho} \left(-\int_\rho^c v_r'(z) dz + \int_c^{1-\rho} v_r'(z) dz\right)
$$
$$
= \sup_{\rho \leq c \leq 1-\rho} (v(\rho) + v(1-\rho) - 2v(c)).
$$

The function $v(\cdot)$ clearly satisfies $|v(z)| \leq 1$ because $f \in \mathcal{C}([0,1]^d, 1)$. This implies that $\int_\rho^{1-\rho} |v_r'(z)| dz \leq 4$. The identity (17) therefore gives

$$
\int_{[\rho, 1-\rho]^d} |f'(x; e_1)| dx
$$
$$
= \int_{(x_2, \ldots, x_d) \in [\rho, 1-\rho]^{d-1}} \left(\int_\rho^{1-\rho} |v_r'(z)| dz\right) dx_2 \ldots dx_d \leq 4.
$$

Similarly, by working with left derivatives of v as opposed to right, we can prove that

$$
\int_{[\rho, 1-\rho]^d} |f'(x; -e_1)| dx \leq 4.
$$

Therefore, the integral $\int_{[\rho, 1-\rho]^d} |m_f|$ is at most $8d$ because it is less than or equal to

$$\sum_{i=1}^{d} \left(\int_{[\rho, 1-\rho]^d} |f'(x; e_i)| dx + \int_{[\rho, 1-\rho]^d} |f'(x; -e_i)| dx \right).$$

This completes the proof of Lemma 4.2. ∎

Remark 4.1: Lemma 4.2 is not true if L_1 is replaced by L_p, for $p > 1$. Indeed, if $d = 1$ and $f_\alpha(x) := \max(0, 1 - (x/\alpha))$ for $0 < \alpha \le 1$ and $g(x) := 0$ for all $x \in [0, 1]$, then it can be easily checked that for $1 \le p < \infty$

$$\|f_\alpha - g\|_p = \frac{\alpha^{1/p}}{(1+p)^{1/p}} \text{ and } \ell_H(V_{f_\alpha}(1), V_g(1)) = \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

As α can be arbitrarily close to zero, this clearly rules out any inequality of the form (14) with the L_1-metric replaced by L_p, for $1 < p \le \infty$.

Remark 4.2: Lemma 4.2 and Bronshtein's result (12) can be used to give an alternative proof of Theorem 3.1 for the special case $p = 1$. Indeed, the scaling identity (1) lets us take $a = 0$, $b = 1$, and $B = 1$. Inequality (14) implies that the covering number $M\left(\mathcal{C}([0,1]^d, 1), \epsilon; L_1\right)$ is less than or equal to

$$M\left(\mathcal{K}^{d+1}(\sqrt{d+1}), \frac{\epsilon}{2(1+20d)}; \ell_H\right).$$

Thus from (12), we deduce the existence of two positive constants c and ϵ_0, depending only on d, such that

$$\log M\left(\mathcal{C}([0,1]^d, 1), \epsilon; L_1\right) \le c\epsilon^{-d/2}$$

whenever $\epsilon \le \epsilon_0$. Note that, by Remark 4.1, this method of proof does not work in the case of L_p, for $1 < p < \infty$.

REFERENCES

[1] A. N. Kolmogorov and V. M. Tihomirov, "ϵ-entropy and ϵ-capacity of sets in function spaces," *Amer. Math. Soc. Transl. (2)*, vol. 17, pp. 277–364, 1961.

[2] D. Dryanov, "Kolmogorov entropy for classes of convex functions," *Constructive Approx.*, vol. 30, pp. 137–153, 2009.

[3] E. M. Bronshtein, "ϵ-entropy of convex sets and functions," *Siberian Math. J.*, vol. 17, pp. 393–398, 1976.

[4] R. M. Dudley, "Metric entropy of some classes of sets with differentiable boundaries," *J. Approx. Theory*, vol. 10, pp. 227–236, 1974.

[5] R. M. Dudley, *Uniform Central Limit Theorems*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[6] M. S. Birman and M. Z. Solomjak, "Piecewise-polynomial approximation of functions of the classes w_p," *Math. USSR Sbornik*, vol. 73, pp. 295–317, 1967.

[7] L. Birgé, "Approximation dans les espaces metriques et theorie de l'estimation," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 65, pp. 181–237, 1983.

[8] L. Le Cam, "Convergence of estimates under dimensionality restrictions," *Ann. Statist.*, vol. 1, pp. 38–53, 1973.

[9] Y. Yang and A. Barron, "Information-theoretic determination of minimax rates of convergence," *Ann. Statist.*, vol. 27, pp. 1564–1599, 1999.

[10] A. Guntuboyina, "Lower bounds for the minimax risk using f divergences, and applications," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2386–2399, Apr. 2011.

[11] S. Van de Geer, *Applications of Empirical Process Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[12] L. Birgé and P. Massart, "Rates of convergence for minimum contrast estimators," *Probab. Theory Related Fields*, vol. 97, pp. 113–150, 1993.

[13] E. Seijo and B. Sen, "Nonparametric least squares estimation of a multivariate convex regression function," *Ann. Statist.*, vol. 39, pp. 1633–1657, 2011.

[14] L. A. Hannah and D. Dunson, *Bayesian nonparametric multivariate convex regression*, 2011, submitted for publication.

[15] A. Seregin and J. A. Wellner, "Nonparametric estimation of multivariate convex-transformed densities," *Ann. Statist.*, vol. 38, pp. 3751–3781, 2010.

[16] M. L. Cule, R. J. Samworth, and M. I. Stewart, "Maximum likelihood estimation of a multi-dimensional log-concave density (with discussion)," *J. Royal Statist. Soc., Ser. B*, vol. 72, pp. 545–600, 2010.

[17] L. Dümbgen, R. J. Samworth, and D. Schuhmacher, "Approximation by log-concave distributions with applications to regression," *Ann. Statist.*, vol. 39, pp. 702–730, 2011.

[18] A. Guntuboyina, "Optimal rates of convergence for the estimation of reconstruction of convex bodies from noisy support function measurements," *Ann. Statist.*, 2011, to be published.

[19] R. J. Gardner, M. Kiderlen, and P. Milanfar, "Convergence of algorithms for reconstructing convex bodies and directional measures," *Ann. Statist.*, vol. 34, pp. 1331–1374, 2006.

[20] P. Massart, *Concentration Inequalities and Model Selection*, ser. Lecture Notes in Mathematics. Berlin, Germany: Springer-Verlag, 2007, vol. 1896.

Adityanand Guntuboyina received the B.Stat. and M.Stat. degrees from the Indian Statistical Institute, Kolkata, in 2024 and 2006, respectively, and the Ph.D. degree in statistics from Yale University, New Haven, CT, in 2011. He was a postdoctoral research associate at the Wharton Statistics Department, University of Pennsylvania from July, 2012 to December, 2012. He is currently an Assistant Professor in the Department of Statistics at the University of California, Berkeley.

He is currently working on nonparametric statistical problems involving convexity typified by the estimation of convex sets and functions from various types of noisy measurements. He is also interested in minimax lower bounds, f-divergences, the concentration of measure phenomenon and respondent driven sampling.

Bodhisattva Sen graduated from Indian Statistical Institute, Kolkata, with a B.Stat. degree in 2002. He obtained his Masters (M.Stat.) degree from the same institute in Statistics in 2004 with a specialization in Mathematical Statistics and Probability. He completed his Ph.D. from the University of Michigan, Ann Arbor, in 2008, and joined Columbia University as the Howard Levene Assistant Professor of Statistics. He spent the academic year 2011–2012 at the University of Cambridge, UK, as a Lecturer in Statslab – a sub-department of the Department of Pure mathematics and Mathematical Statistics – and as a teaching Fellow at St. John's College, Cambridge.

His research interests include shape restricted function estimation, the bootstrap, change-point and threshold models, and statistical applications in astronomy and High Energy physics. Dr. Sen is the recipient of the NSF-CAREER award in July 2011 and, the EPSRC Strategic Fund (from University of Cambridge, UK) in September 2011.