Metric Entropy Under Generalized Convexity

Presented in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

with a Major in

Mathematics

in the

College of Graduate Studies

University of Idaho

by

James Cockreham

Major Professor
Fuchang Gao, Ph.D.

Committee
Lyudmylla Barannyk, Ph.D.
Linh Nguyen, Ph.D.
Stephen Sauchi Lee, Ph.D.

Department Administrator
Christopher J. Williams, Ph.D.

May 2016

Authorization to Submit Dissertation

This dissertation of James Cockreham, submitted for the degree of Doctor of Philosophy with a Major in Mathematics and titled "Metric Entropy Under Generalized Convexity," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor:	Fuchang Gao, Ph.D.	Date
Committee Members:	Lyudmyla Barannyk, Ph.D.	Date
	Stephen Sauchi Lee, Ph.D.	Date
	Linh Nguyen, Ph.D.	Date
Department Administrator:	Christopher J. Williams, Ph.D.	Date

Abstract

Convexity plays an essential role in many areas of mathematics, from geometry, analysis and linear algebra to numerous applications in other areas of mathematics such as optimization. It unifies many apparently diverse mathematical phenomena, and is relevant to engineering and the sciences. In practice, however, convexity does not always hold, which raises the need for suitable generalizations of convexity. In this thesis, I will study generalizations of convexity and use metric entropy to give a numerical quantification of the collections of sets and function classes which satisfy these generalized convexity conditions. In particular, I will estimate the metric entropy of the collection of bounded sets in $\mathbb{R}^d$ with positive reach, the metric entropy of an l_q-hull ($0 < q \leq 1$) in an important case, as well as the upper bound for the metric entropy of separately convex functions in $\mathbb{R}^d$.

Acknowledgements

Completing this dissertation has been a long and sometimes difficult journey. I would not have been able to make it without my advisor Fuchang Gao at my side every step of the way. I would also like to acknowledge my colleague Jesse Oldroyd's assistance with many of the countless tasks that have made up my life since beginning graduate school. The staff in the Department of Mathematics have been consistently supportive and patient for which I also must express my gratitude.

Preface

The results in chapter 3 have been extracted in a preprint titled "Metric Entropy of Sets with Positive Reach", to be submitted for publication jointly with Fuchang Gao.

The results in chapter 4 have been developed and submitted for publication with the title "Metric Entropy of ℓ_q-hulls in Banach spaces of type p, $0 < q \leq 1$."

The results in chapter 5 are being extracted into a preprint which will be published at a later date.

Table of Contents

List of Figures

CHAPTER 1

Introduction

Convexity plays an essential role in many areas of mathematics, from geometry to analysis to linear algebra, and has numerous applications in other areas of mathematics such as optimization. It unifies many diverse mathematical phenomena, and is relevant to engineering and the sciences. In this thesis, I study different generalizations of convexity of sets as well as generalizations of convexity for functions.

Convex sets have been defined and studied in many different settings. The most useful definition to this thesis is that given for a real topological vector space. A subset K of a real topological vector space is said to be convex if for any $x, y \in K$ and any $\lambda \in [0,1]$, the point $\lambda x + (1-\lambda)y \in K$. In other words, for any two points $x, y \in K$, the entire line segment from x to y lies in K. A point $x \in K$ which does not lie in any line segment joining two points of K, except possibly on the end point of a line segment, is called an extreme point of K. If K is convex and compact in a locally convex space, then the set of extreme points $\text{Extr}(K)$ of K is non-empty. Furthermore, by the Krein-Milman theorem, K is the closure of the convex hull of its extreme points; i.e., $K = \overline{\text{conv}(\text{Extr}(K))}$, where, for a set $S \in \mathbb{R}^d$, $\text{conv}(S)$ is defined by

$$\text{conv}(S) := \left\{ x = \sum_{i=1}^{n} \alpha_i x_i \;\middle|\; \alpha_i \geq 0, x_i \in S, \sum_{i=1}^{n} \alpha_i = 1, n \geq 1 \right\}, \tag{1.1}$$

It is straight forward to check that the convex hull of any set is a convex set. This furnishes a convenient source of convex sets.

Convex sets possess useful geometric properties. One in particular, is that, for a normed space X, if $K \subset X$ is closed and convex, and $a \notin K$, then there exists a hyperplane that separates a from K. Furthermore, for any x on the boundary of K, there is a hyperplane containing x such that the entire set K lies on one side of the hyperplane. Such a hyperplane is called a supporting hyperplane at x. This property leads to another common way of generating convex a set: taking intersections of half-spaces.

A function f defined on a convex set K is called convex if it satisfies the inequality

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \tag{1.2}$$

holds for all $x, y \in K$ and every $\lambda \in [0,1]$. If the strict inequality

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \tag{1.3}$$

holds for all $x, y \in K$ and every $\lambda \in (0,1)$, then the function is said to be strictly convex.

There are several equivalent characterizations of conditions (1.2) and (1.3). For instance, it is straightforward to check that any real valued function f is convex on

$$\operatorname{dom}(f) \subset X$$

if and only if for any $x \in \operatorname{dom}(f)$ and any vector $v \in X$, the function

$$g(t) := f(x + tv)$$

is a convex function on

$$\{t \in \mathbb{R} : x + tv \in \operatorname{dom}(f)\}.$$

It is also not difficult to see that if $\operatorname{dom}(f) \subset \mathbb{R}^n$ is a convex set, then f is convex if and only if its epigraph defined by

$$\operatorname{epi}(f) := \left\{ (x,t) \in \mathbb{R}^{n+1} \;\middle|\; t \geq f(x), x \in \operatorname{dom}(f) \right\} \tag{1.4}$$

is a convex set in $\mathbb{R}^{n+1}$.

If f is a differentiable function on $\mathrm{dom}(f)$, where $\mathrm{dom}(f) \subset \mathbb{R}^n$ is a convex set, then f is convex if any only if for all $x, y \in \mathrm{dom}(f)$,

$$f(y) - f(x) \geq \nabla f(x)^T (y - x), \tag{1.5}$$

where ∇f is the gradient of f defined by

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T. \tag{1.6}$$

If f is a twice-differentiable function on $\mathrm{dom}(f)$, where $\mathrm{dom}(f) \subset \mathbb{R}^n$ is a convex set, then f is convex if any only if for all $x \in \mathrm{dom}(f)$, the Hessian matrix $H_f(x)$ is positive semi-definite, where

$$H_f(x) := \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}. \tag{1.7}$$

However, for high dimensions, positive definiteness of the Hessian (1.7) can be difficult to check as are the conditions (1.2) and (1.3). Consequently, in practice, one often verifies the convexity of a function by checking if it can be expressed using some basic convex functions and convexity-preserving operators. Some basic convex functions over $\mathbb{R}^n$ include:

- All norms on $\mathbb{R}^n$;
- The maximum function $\max\{x_1, x_2, \ldots, x_n\}$;
- The log-sum-exp function $\log(\sum_{i=1}^n e^{x_i})$;
- The geometric mean function $(\prod_{i=1}^n x_i)^{1/n}$ over the following subset of $\mathbb{R}^n$:

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i > 0, 1 \leq i \leq n\}. \tag{1.8}$$

Operations that preserve convexity include:

- The operation of taking a nonnegative weighted sum of convex functions: that is, a nonnegative weighted sum of convex functions is convex;
- Composition of a convex function with an affine mapping: if f is a convex function on $\mathrm{dom}(f) \subset \mathbb{R}^n$, A is an $n \times m$ matrix and $b \in \mathbb{R}^n$, then the function

$$g(x) = f(Ax + b) \tag{1.9}$$

  is a convex function on $\{x \in \mathbb{R}^m \mid Ax + b \in \mathrm{dom}(f)\}$.
- Taking the maximum of two convex functions: the maximum of two convex functions is convex on the intersection of their domains; that is, if f and g are convex functions on $\mathrm{dom}(f), \mathrm{dom}(g) \subset \mathbb{R}^n$, then the function h defined by

$$h(x) := \max\{f(x), g(x)\} \tag{1.10}$$

  is a convex function on

$$\mathrm{dom}(h) = \mathrm{dom}(f) \bigcap \mathrm{dom}(g).$$

- Composition of a convex function with a convex and nondecreasing scalar function, i.e., if g is convex function on $\mathrm{dom}(g) \subset \mathbb{R}^n$, and h is a nondecreasing convex function on $\mathbb{R}$, then $f(x) = h(g(x))$ is a convex function on $\mathrm{dom}(g)$;
- The minimization of a convex function on a convex set is convex, i.e., if f is convex in (x, y) and $C \subset \mathrm{dom}(f)$ is a non-empty convex set, then the function

$$g(x) = \inf_{y \in C} f(x, y)$$

  is a convex function on

$$\mathrm{dom}(g) = \{x \mid (x, y) \in \mathrm{dom}(f) \text{ for some } x\},$$

provided that $g(x) > -\infty$.

When a function cannot be expressed as a known convex function using convexity preserving operators, checking convexity is typically very difficult.

On the other hand, when a function is convex, it has some nice properties. For example, one variable convex functions are necessarily continuous and have one-sided derivatives. If f is strictly convex on a compact convex domain, then it has a unique minimizer. Consequently, as a class, the set of convex functions on a fixed domain is much smaller than, say, the class of continuous functions on the same domain. Similarly, the collection of convex sets contained in a bounded region, say $[0,1]^d$, is much smaller than, for instance, the collection of simply connected sets contained in $[0,1]^d$. One of the most effective way to gauge the "size" of a collection of sets, or the "massiveness" of a function class, is to use the metric entropy of the set of functions class, which will be introduced in detail in the next chapter. Roughly speaking, if T is a set in a metric space (X,ρ), then the size of the set T can be gauged by the the quantity

$$\log N(\varepsilon, \mathcal{C}, \rho)$$

with varying ε, where $N(\varepsilon, \mathcal{C}, \rho)$ is the minimum cardinality of ε-net of T, that is,

$$N(\varepsilon, \mathcal{C}, \rho) := \min \left\{ n \mid \exists x_1, \ldots, x_n \in X \text{ such that } \cup_{i=1}^{n} \mathcal{B}(x_i; \varepsilon) \supset \mathcal{C} \right\}, \tag{1.11}$$

where

$$B(x;\varepsilon) = \{ y \in T \mid \rho(x,y) \leq \varepsilon \}.$$

In order to be able to compute metric entropy for classes of sets, we require an appropriate metric to measure the distance between two sets. In this dissertation we employ the most commonly used metric, the Hausdorff distance, $h(\cdot,\cdot)$. Given two bounded sets K and L in $\mathbb{R}^n$ the Hausdorff distance, which we denote by $h(K,L)$, is defined by

$$h(K,L) := \max \left\{ \sup_{k \in K} \inf_{l \in L} \|k - l\|, \sup_{l \in L} \inf_{k \in K} \|k - l\| \right\}. \tag{1.12}$$

In practice, one usually employs the following, equivalent formulation:

$$h(K,L) := \inf\{r \in (0,\infty) \mid K \subset L + B(r), L \subset K + B(r)\}. \tag{1.13}$$

where

$$B(r) = B(0;r) = \{y \in \mathbb{R}^n \mid \|y\| \le r\}.$$

We establish the notation, which will be in use throughout this thesis, $A \asymp B$ which means that there exist two constants $0 < C_1 < C_2 < \infty$ such that $C_1 B \le A \le C_2 B$, then, the following celebrated result of Bronshtein [3] quantifies the "size" of the collection of convex sets contained in $B(1)$, the unit ball in $\mathbb{R}^d$, $d \ge 2$ centered at the origin.

Theorem 1. *(Bronshtein [3]) Let $\mathcal{C}$ be the collection of convex sets contained in $[0,1]^d$, $d \ge 2$. Then for any $0 < \varepsilon \le 1$*

$$\log N(\varepsilon, \mathcal{C}, h) \asymp \varepsilon^{-(d-1)/2},$$

Similarly, if we let Ω be a measurable subset of $\mathbb{R}^d$, and denote

$$\|f - g\|_p = \left(\int_\Omega |f(x) - g(x)|^p dx\right)^{1/p},$$

where $1 \le p < \infty$, then following results proved by Gao and Wellner [15] quantify the massiveness of the collection of convex functions over $\mathbb{R}^d$:

Theorem 2. *(Gao and Wellner [15]) Let Ω be a bounded and closed convex domain in $\mathbb{R}^d$ with non-empty interior. Let $\mathcal{C}_r(\Omega)$, $1 < r \le \infty$, be the set of convex functions on Ω whose $L^r(\Omega)$-norms are bounded by 1. Denote by $|\Omega|$ the volume of Ω.*

(i) There exist constants $c_0 > 0$ depending only on d, such that

$$\log N(\varepsilon, \mathcal{C}_r(\Omega), \|\cdot\|_{L^p(\Omega)}) \ge c_0(|\Omega|^{\frac{1}{p}-\frac{1}{r}}\varepsilon^{-1})^{d/2}.$$

(ii) If Ω has finitely many extreme points, then for any $1 \leq p < r$, there exists a constant C_1 depending on p, d, r such that for any $\varepsilon > 0$,

$$\log N(\varepsilon, \mathcal{C}_r(\Omega), \|\cdot\|_{L^p(\Omega)}) \leq C_1 \varepsilon^{-d/2};$$
$$\log N_{[\,]}(\varepsilon, \mathcal{C}_\infty(\Omega), \|\cdot\|_{L^p(\Omega)}) \leq C_1 |\Omega|^{\frac{d}{2p}} \varepsilon^{-d/2},$$

When $r = \infty$, the same inequality holds for the bracketing entropy $N_{[\,]}(\varepsilon, \mathcal{C}_\infty(\Omega), \|\cdot\|_{L^p}(\Omega))$, which was introduced in definition 3.

In view of the fact that a general compact convex set in $\mathbb{R}^d$ can be approximated by convex sets with finitely many extreme points, one might guess that the rate $\varepsilon^{-d/2}$ holds for a general compact convex set in $\mathbb{R}^d$. This, however, is not the case.

Theorem 3. *(Gao and Wellner [15]) If D is closed unit ball in $\mathbb{R}^d$, then there exists a constant c_2 dependent only on d and p such that for all $0 < \varepsilon < 1$,*

$$\log N(\varepsilon, \mathcal{C}_\infty(D), \|\cdot\|_{L^p(D)}) \geq c_2 \varepsilon^{-\beta},$$

where $\beta = \max\{(d-1)p/2, d/2\}$.

Theorem 1 of Bronshtein is the theoretic foundation of many shape-reconstruction algorithms for high-dimensional convex sets; while Theorem 2 and Theorem 3 play key roles in the study of risk bounds of multivariate convex regression [20]. In these applications, one would often like to see extensions of these results to a more general collections of sets or classes of functions. For example, in image reconstruction of brain activities, it is not realistic to assume that the active regions are convex. Rather, an active region typically consists of several subregions, each of which are *approximately convex*, such as a kidney-shaped region.

Motivated by these applications, the work in this thesis generalizes convexity. There are several different approaches to generalizing convexity for collections of sets. In this thesis, we will focus on the following two approaches:

1. Positive Reach

As we discussed above, a convex set in $\mathbb{R}^d$ can be formed by intersecting half-spaces, and each point on its boundary has a supporting hyperplane. Thus, one can move a half-space (or a ball of infinite radius) along the boundary of a convex set without intersecting the interior of the convex set. Consider, somewhat more generally, sets with the property that one can roll a ball of radius at most c along the boundary of the set. This property was first studied by Federer in [11] under the name "reach".

More precisely, the reach of a set $K \subset \mathbb{R}^d$ is defined as

$$\text{reach}(K) := \sup \{ r \geq 0 \mid \forall y \text{ with } \text{dist}(y, K) \leq r, \exists! x \in K \text{ nearest to } y \}.$$

We say that a set K has positive reach, if $\text{reach}(K) > 0$. Considering its intuitive motivation, positive reach serves as a useful generalization of convexity. Reach is considered in depth in chapter 3.

2. *l^q-Hull,* $0 < q \leq 1$

By the Krein-Milman theorem, every closed convex set in $\mathbb{R}^d$ is the closed convex hull of the set of its extreme points. Thus, every closed convex set is the closure of of a convex hull. If we replace the definition of convex hull given in equation (1.1) by

$$\text{hull}_q(S) := \left\{ x = \sum_{i=1}^{n} \alpha_i x_i \;\middle|\; \alpha_i \geq 0, x_i \in S, \sum_{i=1}^{n} \alpha_i^q = 1, n \geq 1 \right\}, \tag{1.14}$$

where $0 < q \leq 1$, then clearly $\text{hull}_1(S) = \text{conv}(S)$. Thus, $\text{hull}_q(S)$ can be viewed as a generalization of $\text{conv}(S)$. When $0 < q < 1$, the set $\text{hull}_q(S)$ is usually not convex, but has some properties similar to a convex set. We will also consider sets of the form

$$\text{H}_q(S) := \left\{ x = \sum_{i=1}^{n} \alpha_i x_i \;\middle|\; x_i \in S, \sum_{i=1}^{n} |\alpha_i|^q \leq 1, n \geq 1 \right\}, \tag{1.15}$$

These sets are motivated by applications in sparse approximation. Note that when $q = 1$, $\text{H}_q(S)$ is just the absolute convex hull of S; when q is close to 0, then the set $\text{H}_q(S)$ is close to the symmetric star-convex set $\{tx \mid x \in S, |t| \leq 1\}$ with vantage point 0. Thus, for $0 < q < 1$, $H_q(S)$ can be viewed as a continuous deformation from a convex set to a star-convex set.

3. Separate-Con Functions

Just as for convex sets, there are also many different fruitful generalizations of convexixty for classes of functions in $\mathbb{R}^d$. Among the most common generalizations used in practice are polyconvexity, quasiconvex and rank-one convex functions. In this thesis, we consider a condition which is easy to verify. Let a set $D \subset \mathbb{R}^d$ have the property that the intersection of D with any line parallel to a coordinate axis is convex. We say a function f on $\mathrm{dom}(f) = D$ is separately convex if f is a convex function of each single variable while holding the other variables fixed. We say a function f on $\mathrm{dom}(f) = D$ is separately concave if f is a concave function of each single variable while holding the other variables fixed. We say a function f on $\mathrm{dom}(f) = D$ is separate-con if for each variable it is either a concave or convex function of that variable while the other variables are held fixed.

The goal of this thesis is to quantitatively gauge the massiveness of sets and function classes satisfying these three generalizations of convexity. The main results of the thesis are the following:

Theorem 4. *Let $\mathcal{P}_m$ be the collection of d-dimensional sets contained in $B(R)$ that can be expressed as the union of at most m sets of reach greater than c. Then there exists a constant $\Lambda(c,d)$ depending only on c and d, such that for every $0 < \varepsilon < 1$,*

$$\log N(\varepsilon, \mathcal{P}_m, h) \le \Lambda(c,d) m R^d \varepsilon^{-(d-1)/2}.$$

Theorem 5. *If K is a precompact set in a Hilbert space satisfying*

$$\log N(\delta, K, \|\cdot\|) = O(\delta^{-\alpha}),$$

for $\alpha > \frac{2q}{2-q}$, and $0 < q \le 1$, then

$$\log N(\varepsilon, H_q(K), \|\cdot\|) = O\left(\varepsilon^{-\alpha}\right).$$

Theorem 6. *Let $\mathcal{F}([0,1]^d)$ denote the collection of separately convex functions on $[0,1]^d$ that are bounded by 1. Then, for any $1 \le p < \infty$, there exists a constant $C(d,p)$ depending only on d and p, such that for all $0 < \varepsilon < 1$,*

$$\log N(\varepsilon, \mathcal{F}([0,1]^d), \|\cdot\|_p) \le C(d,p)\varepsilon^{-d+\frac{1}{2}}. \tag{1.16}$$

The thesis is organized as follows: In Chapter 2, I collect basic concepts, definitions and known results that are either used in later chapters, or important for conceptual understanding of the topic. In Chapter 3, I study the "size" of the collection of sets contained in $[0,1]^d$ with positive reach. In particular, Theorem 4 is proved. In Chapter 4, I study the metric entropy of l_q-hull $H_q(S)$, provided that we know the metric entropy of S. In particular, Theorem 5 is proved. In Chapter 5, I study the "massiveness" of the class of bounded separate-con functions on $[0,1]^d$, and prove Theorem 6.

CHAPTER 2

Preliminary Results

This chapter collects those definitions, notations and concepts which are integral to the following chapters. It introduces some of the key concepts of metric entropy, some fundamental relations between the so-called Kolmogorov widths and Gelfand widths, and some important metric entropy estimates which are either used in the later chapters or are otherwise conceptually important.

2.1 DEFINITIONS

Metric entropy was first defined by Kolmogorov in [22]. Following Kolmogorov, we first define the ε-covering number for precompact subsets of normed spaces:

Definition 1. *Let A be a precompact subset of a metric space (X, τ). For $\varepsilon > 0$, we define and denote the ε-covering number of A under the metric τ by*

$$N(\varepsilon, A, \tau) := \min\left\{ n \in \mathbb{N} \;\middle|\; \exists x_1, \cdots, x_n \in X,\ A \subset \bigcup_{k=1}^{n} B(x_k, \varepsilon) \right\}, \tag{2.1}$$

where $B(x_k, \varepsilon)$ denotes the closed ball under τ, with center x_k and radius ε. When there is no confusion we refer to $N(\varepsilon, A, \tau)$ simply as the covering number of A.

The logarithm of the ε-covering number of A is called the ε-metric entropy of A or sometimes just the metric entropy of A. More precisely, we have the following definition:

Definition 2. *Let A be a precompact subset of a metric space (X, τ). For $\varepsilon > 0$, we define the ε-metric entropy of A under τ as the quantity* $\log N(\varepsilon, A, \tau)$. *When there is no confusion, we refer to* $\log N(\varepsilon, A, \tau)$ *simply as the metric entropy of A.*

In statistical applications, extra requirements may be placed on metric entropy. For example, instead of closed balls of of radius ε, brackets may be used. In a normed space of a real-valued functions, brackets are defined as follows:

Definition 3. *Let D be a set, $\mathcal{F}$ a linear space of functions $D \to \mathbb{R}$ equipped with norm $\|\cdot\|$. Let $f, h \in \mathcal{F}$ with $f(x) \leq h(x)$ for all $x \in X$. We define the bracket $[f, h]$ by*

$$[f, h] := \{g \in \mathcal{F} \mid f(x) \leq g(x) \leq h(x), \forall x \in X\}. \tag{2.2}$$

If $\|f - h\| \leq \varepsilon$ for all $x \in X$, we say that $[f, h]$ is an ε-bracket.

Using the notion of a bracket, we can now define the ε-bracketing covering number of a precompact subset A of a normed space $(X, \|\cdot\|)$.

Definition 4. *Let $\mathcal{F}$ be a linear space of functions from a set D to $\mathbb{R}$, equipped with norm $\|\cdot\|$. Let A be a subset of $\mathcal{F}$ which is precompact under the topology induced by $\|\cdot\|$. We define the bracketing covering number of A as the minimum number of ε-brackets needed to cover A, that is,*

$$N_{[\,]}(\varepsilon, A, \|\cdot\|) := \min \left\{ n \in \mathbb{N} \;\middle|\; \exists f_1, h_1, \cdots, f_n, h_n \in X,\ A \subset \bigcup_{k=1}^{n} [f_k, h_k] \right\}, \tag{2.3}$$

where, for any i, $[f_i, h_i]$ is an ε-bracket.

Now, we can define bracketing entropy as follows:

Definition 5. *We define the bracketing entropy of $A \subset \mathcal{F}$ to be simply*

$$\log N_{[\,]}(\varepsilon, A, \|\cdot\|). \tag{2.4}$$

To distinguish it from bracketing entropy, definition 2 is also sometimes referred to as entropy without bracketing. The relation between metric entropy and bracketing entropy will be discussed in proposition 7.

It will be convenient to define metric entropy using the concept of an ε-net.

Definition 6. *Given a set A in a metric space (X, τ), a finite set $\mathcal{N}$ is said to be an ε-net of A if, for any $a \in A$, there exists $b \in \mathcal{N}$, such that $\tau(a, b) \leq \varepsilon$.*

Using the notion of an ε-net, we now give an alternate definition of covering number mathematically equivalent to definition (2.1) as:

Definition 7. *The ε-covering number of A under τ is defined as the least cardinality of any ε-net of A.*

Closely related with metric entropy are the so-called entropy numbers, which are defined as follows:

Definition 8. *For a fixed precompact set K in a space equipped with a norm $\|\cdot\|$, the entropy numbers $\varepsilon_l(K)$ are defined by*

$$\varepsilon_l(K) := \inf\{\varepsilon > 0 \mid N(\varepsilon, K, \|\cdot\|) \leq l\}, \tag{2.5}$$

and the dyadic entropy numbers e_l by

$$e_l(K) := \varepsilon_{2^{l-1}}(K) \tag{2.6}$$

If the metric entropy is thought of as a function of ε, then the entropy numbers may be roughly thought of as the corresponding inverse function.

Metric entropy is also fundamentally related to a number of other quantities, among which we mention Kolomogorov width ([21]) and Gelfand width (see [27]), defined as follows:

Definition 9. *Let A be a subset of a normed space $(X, \|\cdot\|)$. We define the Kolmogorov width of A in X by*

$$d_n(A, X) := \inf_{L} \sup_{a \in A} \inf_{l \in L} \|a - l\|, \tag{2.7}$$

where L runs over all n dimensional subspaces of X. Where there is no chance of ambiguity, we will omit mention of the normed space X and simply write $d_n(A)$ instead of $d_n(A, X)$.

The other notion of width which is important for metric entropy is the Gelfand width of a subset of a normed vector space.

Definition 10. *Let A be a subset of a normed space $(X, \|\cdot\|)$. We define the Gelfand width of A by*

$$d^n(A, X) := \inf \left\{ \sup_{a \in L \bigcap A} \|a\| \right\}, \tag{2.8}$$

where the infimum ranges over all n-codimensional subspaces L of X. Where there is no ambiguity, we will omit mention of the normed space X and simply write $d^n(A)$ instead of $d^n(A, X)$.

2.2 SOME IMPORTANT CONNECTIONS

The following well-known proposition summarizes the connection between metric entropy and bracketing entropy:

Proposition 7. *Let A be a precompact subset of some real-valued function space $\mathcal{F}$ equipped with norm $\|\cdot\|$. For any $\varepsilon > 0$,*

$$N(\varepsilon, A, \|\cdot\|) \leq N_{[\,]}(2\varepsilon, A, \|\cdot\|). \tag{2.9}$$

If $\|\cdot\|$ is the supremum norm, then

$$N(\varepsilon, A, \|\cdot\|_\infty) = N_{[\,]}(2\varepsilon, A, \|\cdot\|_\infty). \tag{2.10}$$

Proof. Let

$$N := N_{[\,]}(2\varepsilon, A, \|\cdot\|).$$

Assume without loss of generality that $N < \infty$, for otherwise relation (2.9) is trivial. Then, there exist 2ε-brackets $[f_1, h_1], \cdots, [f_N, h_N]$ such that

$$A \subseteq \bigcup_{k=1}^{N} [f_N, h_N].$$

Each bracket $[f_i, h_i]$ is a closed ball of radius ε with center at $(f_i + h_i)/2$ under the norm $\|\cdot\|$. Thus, the set $\{(f_i + h_i)/2\}_{i=1}^N$ is an ε-net of A. This proves inequality (2.9).

To prove equation (2.10), we notice that if $m := N(\varepsilon, A, \|\cdot\|_\infty)$. Then A can be covered by m closed balls of radius ε, say $B_1, B_2, \dots, B_m$. For each $i = 1, \dots, m$, we define functions f_i and h_i by

$$f_i(x) = \inf\{f(x) : f \in B_i\},$$

and

$$h_i(x) = \sup\{f(x) : f \in B_i\}.$$

Then the bracket $[f_i, h_i]$ is a 2ε-bracket that contains B_i. Thus, $\cup_{i=1}^m [f_i, h_i] \supset A$. Hence $N_{[\,]}(2\varepsilon, A, \|\cdot\|) \leq m$. Together with (2.9), we obtain equation (2.10). □

The connection between Kolmogorov width and metric entropy is less trivial. Here we list two noteworthy relations:

Proposition 8. *([10], the appendix by Levin and Tikhomirov) If*

$$d_n(K) \leq Ce^{-rn}, \qquad \forall n \in \mathbb{N}, \tag{2.11}$$

then

$$\log N(\varepsilon, K, \|\cdot\|) \leq C'r \left(\log \frac{1}{\varepsilon}\right)^2, \qquad \forall \varepsilon \in (0, 1], \tag{2.12}$$

where C' is a constant which does not depend on r or n.

Proposition 9. *(Carl [4]) If for all $n \in \mathbb{N}$*

$$d_{n-1}(K) \leq \frac{C}{n^\alpha}, \tag{2.13}$$

then for all $0 < \varepsilon < 1$,

$$\log N(\varepsilon, K, \|\cdot\|) \leq C'\varepsilon^{-1/\alpha}. \tag{2.14}$$

In addition to relations between widths and metric entropy, there are also some nice duality relations between Kolmogorov widths and Gelfand widths. In particular

Proposition 10. *(Pinkus [27]) Let* $1 \leq p_1, q_1, p_2, q_2 < \infty$ *satisfy*

$$\frac{1}{p_1} + \frac{1}{q_1} = 1, \text{ and} \tag{2.15}$$

$$\frac{1}{p_2} + \frac{1}{q_2} = 1. \tag{2.16}$$

i.e., (p_1, q_1) *and* (p_2, q_2) *are pairs of Hölder conjugates to each other. Then*

$$d^n(B^n_{p_1}, \ell^n_{p_2}) = d_n(B^n_{q_1}, \ell^n_{q_2}). \tag{2.17}$$

In the more general case, one also has the following:

Proposition 11. *(Pinkus [27]) Let* $(X, \|\cdot\|)$ *and* $(Y, |\cdot|)$ *be normed spaces, and let* $K(X, Y)$ *be the collection of compact linear operators between* X *and* Y*. For* $T \in K(X, Y)$*,* T' *denotes the adjoint of* T*. Then,*

$$d_n(T) = d^n(T'), \tag{2.18}$$

where $d_n(T)$ *means* $d_n(T(X), Y)$ *and likewise for* $d^n(T)$*.*

2.3 CONVEX HULLS

There are some important results about the metric entropy of a set and the metric entropy of its convex hull. Because the research in this thesis is partly motivated by these results, we devote special attention to them.

Let K be a precompact subset of a Banach space $(X, \|\cdot\|_X)$, and let $\mathrm{conv}(K)$ be the closed convex hull of K. Suppose we already know the rate of $\log N(\varepsilon, K, \|\cdot\|_X)$ for all $0 < \varepsilon < 1$. Then, some upper bound for the rate of $\log N(\varepsilon, \mathrm{conv}(K), \|\cdot\|_X)$ can be obtained for $0 < \varepsilon < 1$. Among the known results, we list the following two results for precompact sets in a Hilbert space:

Proposition 12. *(Carl, Kyrezi and Pajor [6]) Let K be a precompact subset of a Hilbert space H. For all* $0 < \varepsilon < 1$,

$$\sqrt{\log N(2\varepsilon, \mathrm{conv}(K), \|\cdot\|)} = O\left(\frac{1}{\varepsilon}\int_{\varepsilon/2}^{\mathrm{diam}(K)} \sqrt{\log N(r, K, \|\cdot\|)}dr\right). \tag{2.19}$$

Proposition 13. *(Gao [12]) For all* $0 < \varepsilon < 1$,

$$\log N(2\varepsilon, \mathrm{conv}(K), \|\cdot\|) = O\left(\inf_{\eta}\left\{\frac{\eta^2}{\varepsilon^2} + I^{-1}(\eta)\right\}\right), \tag{2.20}$$

where I^{-1} *is the inverse function of*

$$I(x) = \int_0^x \sqrt{\log N(r, K, \|\cdot\|)}dr.$$

Note that, inequality (2.19) is optimal when $\int_0^{\mathrm{diam}(K)} \sqrt{\log N(r, K, \|\cdot\|)}dr$ diverges; while inequality (2.20) is optimal when $\int_0^{\mathrm{diam}(K)} \sqrt{\log N(r, K, \|\cdot\|)}dr$ converges; see e.g. [14].

The relation between the Gelfand width of a precompact set and that of its absolute convex hull has also been studied. For example, it is shown in [5] that:

Proposition 14. *(Carl, Hinrichs and Pajor [5]) Let K be a precompact in a Hilbert space. Then,*

$$\sqrt{n}d^n(\mathrm{aco}(K)) \leq C\left(1 + \sum_{k=1}^{n}\sqrt{k}e_k(K)\right), \qquad n \in \mathbb{N}, \tag{2.21}$$

where $\mathrm{aco}(K)$ *is the absolute convex hull of K, defined by*

$$\mathrm{aco}(K) := \left\{\sum_{k=1}^{n} c_i k_i \;\middle|\; n \in \mathbb{N},\ |c_1| + \cdots + |c_n| \leq 1, \forall\, i\ k_i \in K\right\} \tag{2.22}$$

and C *is an absolute constant.*

An inequality bounding e_k in terms of d^k is also contained in [5]:

Proposition 15. *(Carl, Hinrichs and Pajor [5]) Let K be a precompact in a Hilbert space. Then,*

$$\sum_{k=1}^{n} k^{s/r-1}(e_k(\mathrm{conv}(K)))^s \leq C(r,s) \sum_{k=1}^{n} k^{s/r-1}(d^k(\mathrm{conv}(K)))^s, \tag{2.23}$$

where $0 < s, r \leq \infty$ and $C(r,s)$ is a constant depending only on r and s.

These propositions are especially useful in function spaces. For example, if K is the set of indicator functions of the intervals $[0,a]$, $0 < a \leq 1$. Then $\mathrm{conv}(K)$ is the set of all non-negative monotonic functions on $[0,1]$ that is bounded by 1. This greatly simplifies the problem. These two propositions also motivated us to study so-called l_q-hulls for $0 < q \leq 1$ in the next chapter.

2.4 SOME METRIC ENTROPY ESTIMATES

In this section we collect some well-known metric entropy estimates. Some of which will serve as basic facts and will be used in what follows; others are included here to give readers an impression about the typical growth rate of metric entropy estimates and how the growth rate captures the information about the "size" of the set (or function class).

2.4.1 *Ellipsoidal Classes*

The following known result is a basic fact, and will be used later.

Proposition 16. *([9], page 98) Let $B_{\ell_p^n}$ be the unit ball of ℓ_p^n, $0 \leq p < q < \infty$. Then the ε-metric entropy of $B_{\ell_p^n}$, under the ℓ_q norm is given by*

$$\log N(\varepsilon, B_{\ell_p^n}, \ell_q^n) \asymp \begin{cases} \frac{1}{\varepsilon^s}\log(2n\varepsilon^s), & n\varepsilon^s > 1 \\ n\log(\frac{2}{n\varepsilon^s}), & n\varepsilon^s < 1 \end{cases}, \tag{2.24}$$

where we set $s := \frac{pq}{q-p}$.

To illustrate the usefulness of this result, we introduce two estimates which appear in applications.

Corollary 17. *(Kolmogorov and Tikhomirov [24]) Let* $\phi_1, \cdots, \phi_k, \cdots$ *be an orthonormal basis for* $L^2([0,1])$*, and let* $\{b_k\}$ *be a sequence of constants where* $b_k \asymp k^\alpha$*, and* $\alpha > 0$*, so* $b_k \to \infty$*. We define the ellipsoidal class* $\mathcal{E}(\{b_k\}, C)$ *by*

$$\mathcal{E}(\{b_k\}, C) := \left\{ g = \sum_{k=1}^{\infty} \xi_k \phi_k \,\middle|\, \sum_{k=1}^{\infty} \xi_k^2 b_k^2 < C \right\}. \tag{2.25}$$

Then, the class $\mathcal{E}(\{b_k\}, C)$ *has metric entropy under the* L^2 *norm of*

$$\log N(\varepsilon, \mathcal{E}(\{b_k\}, C), \|\cdot\|_{L^2}) \asymp \varepsilon^{-1/\alpha}. \tag{2.26}$$

With the assumption that $b_k \asymp k^\alpha$ removed, much is still known about these classes. For example, Mitjagin [26] studied cross-sections of $\mathcal{E}(\{b_k\}, C)$, i.e., functions of the form

$$\sum_{k=1}^{m} \xi_k \phi_k, \tag{2.27}$$

and applying results similar to that from equation (4.2) showed that, in general, these ellipsoidal classes have entropy satisfying

$$m\left(\frac{2}{\varepsilon}\right) \log \frac{8}{\varepsilon} \geq \log N(\varepsilon, \mathcal{C}(\{b_k\}, C), \|\cdot\|) \geq m\left(\frac{1}{2\varepsilon e}\right) \tag{2.28}$$

where $e \approx 2.71828$ denotes the base of the natural logarithm.

Similarly, the following estimate can be obtained:

Corollary 18. *(Smoljak [28]) Let* C *be a constant,* $k \geq 0$ *and* $\alpha > \frac{1}{2}$*. Define* $E_d^{k,\alpha}(C)$ *as the set of all functions*

$$g(x_1, \cdots, x_d) = \sum_{m_1, \cdots, m_d = -\infty}^{\infty} \left(a_{m_1, \cdots, a_{m_d}} \cos\left(\sum_{i=1}^{d} 2\pi m_i x_i\right) + b_{m_1, \cdots, m_d} \sin\left(\sum_{i=1}^{d} 2\pi m_i x_i\right)\right) \tag{2.29}$$

over $[0,1]^d$ *subject to the conditions*

$$\sqrt{a^2_{m_1,\cdots,m_d} + b^2_{m_1,\cdots,m_d}} \leq C(\overline{m}_1 \cdots \overline{m}_d)^{-\alpha}(\log^k(\overline{m}_1 \cdots \overline{m}_d) + 1) \tag{2.30}$$

for all $m_1, \cdots, m_d \in \mathbb{R}$, *with* $\overline{m} = \max\{m, 1\}$. *Then, under the* $\|\cdot\|_{L_2}$ *norm,*

$$\log N(\varepsilon, E_d^{k,\alpha}, \|\cdot\|_{L_2}) \asymp \left(\frac{1}{\varepsilon}\right)^{1/(\alpha-1/2)}. \tag{2.31}$$

In analogy with the definition of $E_d^{k,\alpha}(C)$, one may define a class of functions, $G_d^{k,\alpha}(C)$, of the form given in equation (2.29), but satisfying the constraint

$$\sum_{m_1,\cdots,m_d=-\infty}^{\infty} (\overline{m}_1 \cdots \overline{m}_d)^{2\alpha} \left(a^2_{m_1,\cdots,m_d} + b^2_{m_1,\cdots,m_d}\right) \leq C^2 \tag{2.32}$$

instead of the constraint given in (2.30). Then, $G_d^{k,\alpha}(C)$ has metric entropy (see [28]):

$$\log N(\varepsilon, G_d^{k,\alpha}(C), \|\cdot\|_{L_2}) \asymp \left(\frac{1}{\varepsilon}\right)^{1/\alpha} \log^{d-1}\left(\frac{1}{\varepsilon}\right). \tag{2.33}$$

2.4.2 *Finite Differences*

We will use results from this section as a starting point for the research in this dissertation, especially chapter 5.

For a function $g\colon \mathbb{R} \to \mathbb{R}$, define the finite difference operator $\Delta_h g$ by $\Delta_h g(x) = g(x+h) - g(x)$. It follows from direct computation that the k-th such difference, denoted Δ_h^k is given by

$$\Delta_h^k g(x) := \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} g(x+jh). \tag{2.34}$$

Let C be a constant. Let $q \geq 1$ and $\alpha \geq (\frac{1}{q} - \frac{1}{2})^+$, where $a^+ = \max\{a, 0\}$. Define the class of Lipschitz functions by

$$Lip_{\alpha,q}(C) := \{g\colon [0,1] \to [-C, C] \mid \|\Delta_h g(x)\|_{L^q} \leq Ch^\alpha\}. \tag{2.35}$$

Then, the metric entropy of $Lip_{\alpha,q}(C)$ is:

Proposition 19. *(Birman and Solomjak[1]) The metric entropy of $Lip_{\alpha,q}(C)$ is given by*

$$\log N(\varepsilon, Lip_{\alpha,q}(C), \|\cdot\|_{L^2}) \asymp \varepsilon^{-1/\alpha}. \tag{2.36}$$

Define the class of functions of bounded variation by

$$BV(C) := \{g\colon [0,1] \to [-C, C] \mid V(g) \leq C\} \tag{2.37}$$

where $V(g)$ is the variation of g, defined by:

$$V(g) := \sup \sum_{i=1}^{m} |g(x_{i+1}) - g(x_i)| \tag{2.38}$$

where the supremeum is taken over all partitions $\{x_1, \cdots, x_d\}$ of $[0,1]$. Since (see e.g. [8])

$$Lip_{1,\infty}(C) \subset BV(C) \subset Lip_{1,1}(C), \tag{2.39}$$

we immediately obtain, via Proposition 19, the following estimate of metric entropy for the class of bounded variation functions:

Corollary 20. *For any $\varepsilon > 0$,*

$$\log N(\varepsilon, BV(C), \|\cdot\|_{L^2}) \asymp \frac{1}{\varepsilon}. \tag{2.40}$$

Because a bounded monotonic function on $\mathbb{R}$ is automatically of bounded variation, we have:

Corollary 21. *Let $\mathcal{F}$ denote the set of all monotone increasing functions $f\colon I \to [0,1]$, where $I \subseteq \mathbb{R}$ is an arbitrary bounded interval in $\mathbb{R}$. The entropy of $\mathcal{F}$ is given by:*

$$\log N(\varepsilon, \mathcal{F}, \|\cdot\|_{L^p}) \asymp \frac{1}{\varepsilon} \tag{2.41}$$

Remark 22. *It can be proved directly that*

$$\log N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{L^p}) \asymp \frac{1}{\varepsilon}.$$

(See e.g. [32] or [2] for a simple proof.) Because any function of bounded variation can be written as a difference of two monotone functions (Mcdonald and Weiss [25]), we have $BV(C) \subset \mathcal{F} - \mathcal{F}$. *Thus, the following estimate, which is stronger than that contained in equation (2.40), also holds:*

$$\log N_{[\,]}(\varepsilon, BV(C), \|\cdot\|_{L^2}) \asymp \frac{1}{\varepsilon}.$$

For a function $g\colon \mathbb{R}^n \to \mathbb{R}$, denote the k_i-th difference operator with respect to the variable x_i by $\Delta_h^{k_i}$, that is,

$$\Delta_h^k g(x) := \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} g(x + jhe_i), \tag{2.42}$$

where e_i is the vector whose i-th coordinate is 1 and other coordinates are 0.

The metric entropy of the class of functions with the properties mentioned above is given as follows:

Proposition 23. *(Timan [31], page 279) Let A consist of all the functions $f : [0,1]^n \to [0,1]$ such that for each $1 \le i \le n$, there exists a constant M_i such that for all $0 < t < 1$, $\sup_{0<h\le t} \|\Delta_h^k f\|_2 \le M_i t^{\beta_i}$, where $0 < \beta_i \le k_i$. Then,*

$$\log N(\varepsilon, A, \|\cdot\|_2) \asymp \varepsilon^{-\sum_{i=1}^{n} \beta_i^{-1}}. \tag{2.43}$$

2.4.3 *Differentiable Functions*

Although the results from this subsection and the next section will not be be used in the dissertation, we include them in order to illustrate the rates of some typical metric entropy estimates and how the rates of these estimates are affected by the presence of a particularly strong hypothesis such as analyticity for instance.

To introduce these results, we let $I = [a, b]$, and let $W_p^r(I)$ denote the collection of all functions $f\colon I \to \mathbb{R}$ such that $f^{(r-p)}$ is absolutely continuous and $f^{(r)} \in L^p(A)$. We further define

$$B_p^r := \left\{ f \in W_p^r \;\middle|\; \left\| f^{(r)} \right\|_p \leq 1 \right\}. \tag{2.44}$$

For $0 \leq h \leq b - a$, we define the modulus of continuity $w(\cdot, \cdot)_p$ by

$$w(g, h)_p := \max_{0<t\leq h} \left(\int_a^{b-h} |g(x+t) - g(x)|^p \, dx \right)^{1/p}. \tag{2.45}$$

For $\alpha = r + \beta$ with $0 \leq \beta \leq 1$, we define

$$Lip_p^\alpha := \left\{ f \in W_p^\alpha \;\middle|\; w(f^{(r)}, t)_p \leq t^\beta \right\}. \tag{2.46}$$

If $p = \infty$, we further define $\overline{Lip_\infty^\alpha}$ as follows:

$$\overline{Lip_\infty^\alpha} := \begin{cases} \left\{ f \in Lip_\infty^\alpha \;\middle|\; \left\| f^{(r)} \right\|_\infty \leq 1 \right\} & \text{if } \beta = 0, \\ \left\{ f \in Lip_\infty^\alpha \;\middle|\; \left| f^{(r)}(x) - f^{(r)}(y) \right| \leq |x - y|^\beta \right\} & \text{if } \beta \neq 0. \end{cases} \tag{2.47}$$

With the notations defined above, we have:

Proposition 24. *(Temljakov [29]) There exists a constant C such that for all* $0 < \varepsilon < 1$*,*

$$\log N(\varepsilon, \overline{Lip_p^\alpha}, \|\cdot\|_q) \geq C\varepsilon^{-1/\alpha}. \tag{2.48}$$

If $\alpha \geq \frac{1}{p} - \frac{1}{q}$, then the reverse to inequality (2.48) has been shown via Taylor expansions:

Proposition 25. *(Birman and Solomyak [1]) If* $\alpha \geq \frac{1}{p} - \frac{1}{q}$*, then*

$$\log N(\varepsilon, \overline{Lip_p^\alpha}, \|\cdot\|_q) \asymp \varepsilon^{-1/\alpha}. \tag{2.49}$$

When $p = q = \infty$, the estimate (2.49) has been extended to higher dimensional cases:

Proposition 26. *(Kolmogorov [23]) Under the notations defined in the previous proposition,*

$$\log N(\varepsilon, \overline{B^{\alpha}_{\infty}}([0,1]^d), \|\cdot\|_{\infty}) \asymp \varepsilon^{-d/\alpha}. \tag{2.50}$$

2.4.4 *Analytic Functions*

When the functions are analytic, the metric entropy rate is small, which reflects the intuition that the class of analytic functions must be "small" since analyticity is such a strong condition. Here we list a few typical results:

Proposition 27. *(Vitushkin [33]) We denote the closed, 0 centered ball of radius r by $D(r)$. Let $\mathcal{A}(D(r))$ denote all the analytic functions on an open connected set G which are uniformly bounded by 1.*

$$\log N(\varepsilon, \mathcal{A}(D(r)), \|\cdot\|_{L^\infty(D(r))}) = \frac{1}{\log r} \log\log\frac{1}{\varepsilon} + O(\log\frac{1}{\varepsilon}\log\log\frac{1}{\varepsilon}). \tag{2.51}$$

It is noteworthy that equation (2.51) is exact and gives the non-asymptotic term precisely. For general connected open sets, similar but less precise results can be obtained:

Proposition 28. *(Widom [34]) Let $G \subset \mathbb{C}$ be a connected open set, $K \subset G$ be a compact set, and we let $\mathcal{A}(G)$ denote all the analytic functions on G which are uniformly bounded by 1.*

$$\log N(\varepsilon, \mathcal{A}(G), \|\cdot\|_{L^\infty(K)}) \asymp C(K,G)\log\log\frac{1}{\varepsilon}, \tag{2.52}$$

where $C(K,G)$ is a constant depending only on K and G.

It is shown in ([34]) that this constant is in fact $\frac{1}{V(K,G)}$ where $V(K,G)$ is defined by the equation

$$e^{-V(K,G)} = \lim_{n\to\infty}\left[d_n(\mathcal{A}(G)), \|\cdot\|_{L^\infty(D(r))}\right]^{1/n}. \tag{2.53}$$

In higher dimensions, if $G \subseteq \mathbb{C}^d$ is a bounded, open and connected set, $K \subseteq G$ denoting a compact set with non-empty interior, then it is shown in Kolmogorov [23] that

$$\log N(\varepsilon, \mathcal{A}(G), \|\cdot\|_{L^\infty(K)}) \asymp (\log \frac{1}{\varepsilon})^{d+1}. \tag{2.54}$$

CHAPTER 3

Metric Entropy of Sets With Positive Reach

3.1 Sets of Positive Reach

Convexity is a fundamental part of analysis. Unfortunately, convexity may not hold in practice. To meet the needs of practitioners from diverse areas of mathematics and the sciences, several generalizations have been proposed and studied. Among the most popular is the notion of positive reach. In Federer's seminal paper [11] of 1959, the following definition of the reach of a set is introduced:

Definition 11. *For a set $K \subset \mathbb{R}^d$, the reach of K is defined as*

$$\operatorname{reach}(K) := \sup \{ r \geq 0 \mid \forall y \textit{ with } \operatorname{dist}(y, K) \leq r, \exists! x \in K \textit{ nearest to } y \}.$$

Definition 12. *A set $K \subset \mathbb{R}^d$ is said to have positive reach if* $\operatorname{reach}(K) > 0$.

It is easy to check that if $\operatorname{reach}(K) > 0$, then K is necessarily a closed set. Furthermore, K is closed and convex if and only if $\operatorname{reach}(K) = \infty$. Thus, positive reach is a generalization of convexity.

Let us remark that a set of positive reach can have a very difference appearance from that of a convex set. For example, a set of positive reach could have a cavity. It is easy to check that the complement of any open ball of radius R is a set of positive reach with $\operatorname{reach}(K) = R$. Furthermore, a set of positive reach does not need to be connected. Indeed, any finite set is of positive reach: If K is a finite set, then

$$\operatorname{reach}(K) = \min \left\{ \frac{1}{2} \|x - y\| \;\middle|\; x, y \in K, x \neq y \right\}.$$

For convex sets, it is clear that the intersection of two convex sets is still convex. In terms of reach, we have the following:

$$\operatorname{reach}(K) = \operatorname{reach}(L) = +\infty \;\Rightarrow\; \operatorname{reach}(K \cap L) = +\infty.$$

It is important to remark that the intersection of two sets with positive reach may no longer of positive reach. In fact, we have the following:

Proposition 29. *For any $r > 0$, there exists set K with* $\operatorname{reach}(K) = r$*, and a closed convex set L, i.e.,* $\operatorname{reach}(L) = \infty$*, such that* $\operatorname{reach}(K \cap L) = 0$*.*

Proof. Let

$$K = \left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\| \geq r \right\}.$$

We prove that $\operatorname{reach}(K) = r$. For any $x \in \mathbb{R}^2$, if $\operatorname{dist}(x, K) = 0$, then $x \in K$. So, there exists a unique point in K that is closest to x. If $0 < \operatorname{dist}(x, K) < r$, then $x \neq 0$ and $\operatorname{dist}(x, K) = r - \|x\|$. In this case, the closed ball $B(x, r - \|x\|)$ intersects K at the unique point $\frac{r}{\|x\|} x$. Therefore, $\operatorname{reach}(K) \geq r$.

On the other hand, because $\operatorname{dist}(0, K) = r$, and the closed ball $B(0, r)$ intersects K at infinitely many points (indeed, at every point on $\left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\| = r \right\}$). Hence $\operatorname{reach}(K) \leq r$. Therefore $\operatorname{reach}(K) = r$.

To construct the convex set L we choose a strictly decreasing sequence of real numbers $\{\theta_n\}_{n=1}^{\infty}$ in $(0, \pi)$ that converges to 0, and define

$$x_n = (r \cos(\theta_n), r \sin(\theta_n)) \in \mathbb{R}^2.$$

Thus, the points x_n, $n \geq 1$, are on the upper half of the circle $\left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\| = r \right\}$.

For $n \geq 1$, let L_n be the line containing x_{2n-1} and x_{2n}, and let H_n be the region below the line L_n, including the points on the line L_n. Furthermore, we let L_0 be the line containing x_1 and the point $(r, 0) \in \mathbb{R}^2$, and let H_0 be the region *above* the line L_0, including the points on the line L_0. Thus, H_n is a closed half space for each $n \geq 0$. Consequently, the set

$$L := \bigcap_{n=0}^{\infty} H_n$$

is a closed convex set in $\mathbb{R}^2$. Now, let us look at the intersection $K \cap L$. It is not difficult to see that

$$K \cap L = \bigcup_{n=1}^{\infty} \Delta_n,$$

where Δ_n is the closed region bounded by the line L_n, the line L_{n+1}, and the short arc $x_{2n}x_{2n+1}$. If we let $\xi_n = \frac{1}{2}(x_{2n} + x_{2n+1})$ for $n \geq 2$. Then it is easy to check that

$$\operatorname{dist}(\xi_n, K \cap L) = \frac{1}{2}\|x_{2n+1} - x_{2n}\|.$$

Because there are two points on $K \cap L$, namely, x_{2n} and x_{2n+1}, which are closest to ξ_n, we have

$$\operatorname{reach}(K \cap L) \leq \frac{1}{2}\|x_{2n+1} - x_{2n}\|,$$

for all $n \geq 1$. However,

$$\lim_{n\to\infty} \|x_{2n+1} - x_{2n}\| = 0.$$

Therefore, $\operatorname{reach}(K \cap L) = 0$. □

While Proposition 29 says that in general, the intersection of two sets with positive reach may not have positive reach, the following proposition says that under a certain condition, the intersection can still have positive reach.

Proposition 30. *(Colesanti and Manselli [7], Theorem 3.10) If* $\operatorname{reach}(K) \geq R > 0$*, and* L *is a closed set with the property that whenever* $a, b \in L$*, the set* $I(a, b, R) \subset L$*, where* $I(a, b, R)$ *is the intersection of all closed balls of radius* R *containing* a *and* b*, then* $\operatorname{reach}(K \cap L) \geq R$*. In particular, if* L *is a closed ball of radius no larger than* R*, then* $\operatorname{reach}(K \cap L) \geq R$*.*

On the other hand, sets with positive reach have many nice properties. In particular, Federer [11] extended both the classical Steiner's formula for convex sets, and the fundamental Weyl's Tube Formula for differential manifolds to sets of positive reach. We refer the readers to a nice survey paper [30] for the research on positive reach up to 2009.

Among the many nice properties of sets with positive reach, we recall the following property that was discovered by Federer:

Proposition 31. *((6) of Theorem 4.8 in Federer [11]) For any* $a \in \partial K$*, and any* $v \in \mathbb{R}^d$*, if*

$$0 < \tau = \sup\{t \mid \operatorname{dist}(a + tv, K) = \|tv\|\} < \infty,$$

then $\|\tau v\| \geq \operatorname{reach}(K)$*.*

We use Proposition 31 to prove the following important property of sets with positive reach, which will be used in an essential way in the later proof.

Lemma 32. *If K is a set in $\mathbb{R}^d$ with positive reach, then the following are true:*

i) For every $0 < r < \mathrm{reach}(K)$, and every $y \in \mathbb{R}^d$ with $\mathrm{dist}(y, K) < \mathrm{reach}(K)$, if x is the unique point on K that is closest to y, then, the closed ball $B(z, r)$ intersects K precisely at x, where

$$z = x + \frac{r}{\|y - x\|}(y - x).$$

ii) For every $0 < r < \mathrm{reach}(K)$, every point on the boundary of K is touchable by a closed ball of radius r from outside, that is, for every $0 < r < \mathrm{reach}(K)$ and every $x_0 \in \partial K$, there exists a $w \in \mathbb{R}^d$, such that the closed ball $B(w, r)$ intersects K precisely at x_0.

Proof. We prove i) by contradiction. Suppose i) is false, then there exist a $y \in R^d$ with $0 < \mathrm{dist}(y, K) < \mathrm{reach}(K)$, and an $x \in K$ with $\|y - x\| = \mathrm{dist}(y, K)$, and $0 < r < \mathrm{reach}(K)$ such that

$$\|z - x\| > \mathrm{dist}(z, K),$$

where

$$z = x + \frac{r}{\|y - x\|}(y - x).$$

Then,

$$0 < 1 \leq \tau = \sup\left\{t \mid \mathrm{dist}(x + t(y - x), K) = \|t(y - x)\|\right\} \leq \frac{r}{\|y - x\|} < \infty.$$

By applying Proposition 31, we have

$$r = \left\| \frac{r}{\|y - x\|} \cdot (y - x) \right\| \geq \tau \|y - x\| \geq \mathrm{reach}(K),$$

which is a contradiction. This proves i).

Now, we use i) to prove ii). Because x_0 is on the boundary of K, there exists $N_0 > 0$ such that for every integer $n \geq N_0$, we can choose a point y_n outside K so that

$\|y_n - x_0\| < r/n$. Let x_n be the unique point in K that is closest to y_n. Let

$$w_n = x_n + \frac{r}{\|y_n - x_n\|}(y_n - x_n).$$

By i), the closed ball $B(w_n, r)$ intersects K precisely at x_n, and y_n lies between x_n and w_n. Since

$$\|w_n - x_0\| \leq \|w_n - x_n\| + \|x_n - x_0\| \leq (1 + 1/n)r$$

for all $n \geq N_0$, the sequence $\{w_n\}$ is bounded in $\mathbb{R}^d$. Hence it contains a subsequence that converges to some $w_0 \in \mathbb{R}^d$. Clearly, we have

$$\|w_0 - x_0\| = r = \text{dist}(w_0, K).$$

Since $r < \text{reach}(K)$, the set closed ball $B(w_0, r)$ intersects K precisely at x_0. This finishes the proof of ii). □

3.2 HAUSDORFF DISTANCE

A key concept in measuring the distance between two bounded sets is the so-called Hausdorff distance. Let $B(r)$ be the closed ball in $\mathbb{R}^d$ with center at 0 and radius r, and let K_1 and K_2 be two bounded closed sets in $\mathbb{R}^d$. The *Hausdorff distance* between K_1 and K_2 is defined to be the minimum value of r so that

$$K_1 \subset K_2 + B(r), \tag{3.1}$$

and

$$K_2 \subset K_1 + B(r). \tag{3.2}$$

Hausdorff distance is commonly used in shape approximation, for instance in computer graphics the Hausdorff distance is used to measure the difference between two different representations of the same 3D object [19], particularly when generating level of detail for efficient display of complex 3D models. A celebrated result of

Bronshtein states that if $\mathcal{C}_d$ is the collection of closed convex subsets of $B(1)$, then there exists a constant $\lambda(d)$ depending only on d such that for all $\varepsilon > 0$, $\mathcal{C}_d$ has an ε-net in Hausdörff distance of cardinality

$$N(\varepsilon, \mathcal{C}_d, h) \leq \exp\left(\lambda(d)\varepsilon^{-(d-1)/2}\right).$$

This result provides a theoretic base of approximation algorithms of convex bodies. It also plays a key role in the study of metric entropy of convex functions, see, for instance, [18] and [15]. In applications, one often wants an extension of Bronshtein's result to non-convex bodies. For example, in image reconstruction of brain activities, it is not realistic to assume that the active regions are convex. Rather, an active region typically consists of several subregions, each of which are *approximately convex*, such a kidney-shaped region. Motivated by these applications, in this chapter, we extend Bronshtein's theorem to sets with *positive reach*. We believe that this result, and more broadly, the notion of *reach* itself, will useful to workers in applied fields where certain regularity conditions may not hold in practice.

Having defined reach and Hausdorff distance, we can now state the main result of this chapter:

Theorem 33. *Let $\mathcal{K}$ be the collection of all sets K contained in $B(1)$, with* $\operatorname{reach}(K) > c$*. For any $\varepsilon > 0$,*

$$\log N(\varepsilon, \mathcal{K}, h) \leq \Lambda(c, d)\varepsilon^{-(d-1)/2},$$

where $\Lambda(c, d)$ is a constant depending only on c and d.

3.3 COATING A SET OF POSITIVE REACH

The key idea of our approach to study the metric entropy of $\mathcal{K}$ is to smooth the boundary of each set $K \in \mathcal{K}$ by means of "coating" it with a closed ball of radius $\delta > 0$. To be precise, let K be a bounded closed set in $\mathbb{R}^d$. We define

$$K^r := K + B(r).$$

We call K^r the r-coating of K. We also define

$$\mathcal{K}^r := \{K^r \mid K \in \mathcal{K}\}.$$

The following lemma establishes the equivalence between the metric entropies of $\mathcal{K}$ and $\mathcal{K}^r$.

Lemma 34. *For all* $0 < \delta, r \leq c$,

$$N(\delta, \mathcal{K}^r, h) \leq N(\delta, \mathcal{K}, h) \leq N(\delta/2, \mathcal{K}^r, h).$$

Proof. Let $C, D \in \mathcal{K}$, and let $h(C, D) = \delta$. We first show that

$$h(C + B(r), D + B(r)) = h(C, D). \tag{3.3}$$

We begin by establishing that

$$h(C + B(r), D + B(r)) \leq h(C, D).$$

By definition of Hausdorff distance,

$$C \subset D + B(\delta),$$

and

$$D \subset C + B(\delta).$$

Consequently,

$$C + B(r) \subset D + B(r) + B(\delta)$$

and

$$D + B(r) \subset C + B(r) + B(\delta),$$

which implies that

$$h(C + B(r), D + B(r)) \leq \delta.$$

This is one direction of (3.3).

On the other hand, by the definition of Haüsdorff distance, together with the compactness of C and D, either there exists $x \in C$ such that $\text{dist}(x, D) = \delta$, or there exists $y \in D$ such that $\text{dist}(y, C) = \delta$. Without loss of generality, we assume $\text{dist}(x, D) = \delta$.

Since D is closed, there exists $y \in D$ such that $\|x - y\| = \delta$. That is, y is the point in D that is closest to x.

Since $\text{reach}(D) > \max\{r, \delta\} := r^*$, by Lemma 32, there exists a closed ball $B(w, r^*)$ with center at w and radius r^* such that $B(w, r^*)$ intersects D only at y, and x lies on the line segment from y to w. Let $v = x + \frac{r}{\|w-x\|}(w - x)$. Clearly, $v \in C + B(r)$, and $\text{dist}(v, D + B(r)) = \delta$. Thus,

$$h(C + B(r), D + B(r)) \geq \delta = h(C, D).$$

This finishes the proof of equation (3.3).

We proceed to finish the lemma. We claim first that, if $\{B_i\}$ is an δ-net for K, $\{B_i + B(r)\}$ is an δ-net for K^r. Let $E \in \mathcal{K}^r$. Then, for some $E_1 \in \mathcal{K}$, we have that $E = E_1 + B(r)$. Let B_i be an element of the δ-net with $h(E_1, B_i) \leq \delta$. Then,

$$h(E, B_i + B(r)) = h(E_1 + B(r), B_i + B(r)) \tag{3.4}$$
$$= h(E_1, B_i). \tag{3.5}$$
$$\leq \delta \tag{3.6}$$

Thus, $\{B_i + B(r)\}$ is a δ-net for $\mathcal{K}^r$ as claimed. If, in particular, $\{B_i\}$ is a δ-net for $\mathcal{K}$ with minimum cardinality, this implies that

$$N(\delta, \mathcal{K}^r, h) \leq N(\delta, \mathcal{K}, h).$$

Now we show that

$$N(\delta/2, \mathcal{K}, h) \leq N(\frac{\delta}{2}, \mathcal{K}^r, h).$$

Let $\{C_i\}$ be a δ-net of $\mathcal{K}^r$ of minimum cardinality. For each i, choose K_i in K such that

$$h(C_i, K_i + B(r)) \leq \frac{\delta}{2}. \tag{3.7}$$

Then, for any $A \in \mathcal{K}$ we have for appropriate i:

$$\begin{aligned} h(A, K_i) &= h(A + B(r), K_i + B(r)) && (3.8)\\ &\leq h(A + B(r), C_i + B(r)) + h(C_i + B(r), K_i + B(r)) && (3.9)\\ &\leq \frac{\delta}{2} + \frac{\delta}{2} && (3.10)\\ &= \delta. && (3.11) \end{aligned}$$

Thus,

$$N(\delta, \mathcal{K}, h) \leq N(\delta/2, \mathcal{K}^r, h).$$

This completes the proof. □

3.4 SOME PROPERTIES OF COATED SETS

We now prove some key properties of the coated sets of sets with positive reach. We will use these properties in the later proofs.

The first property is that a certain kind of coated sets are star-convex.

Lemma 35. *Suppose $K \subset B(x_0, r/2)$ with* reach$(K) > r$. *Then, every ray emanating from x_0 intersects the boundary of K^r at exactly one point. That is, K^r is star-convex with vantage point x_0.*

Proof. Suppose a ray emanating from x_0 intersects the boundary K^r at two distinct points y and z, with $\|y - x_0\| < \|z - x_0\|$. Since reach$(K) > r$, there exists a unique $w \in K$ such that $\|z - w\| = r$. Let m be the middle point of y and z, and let H be the hyperplane that passes through m and is perpendicular to the line segment from y to z. Because $\|y - w\| \geq r = \|z - w\|$, the hyperplane H separates w and y, with y on the side containing x_0 and w on the other side. Since $w, x_0 \in B(0, \frac{1}{4}r)$, we have

$$\|w - x_0\| \leq \frac{1}{2}r.$$

On the other hand, because $\|y - w\| \geq r$, we have

$$\|y - x_0\| \geq \|y - w\| - \|x_0 - w\| \geq r - \frac{1}{2}r = \frac{1}{2}r,$$

which implies that

$$\|w - x_0\| \geq \|m - x_0\| > \|y - x_0\| \geq \frac{1}{2}r.$$

This is a contradiction. Hence any ray emanating from x_0 intersects the boundary of K^r only at one point. Consequently, K^r is star-convex. □

The next lemma describes an important property coated sets.

Lemma 36. *Suppose* $K \subset B(x_0, r/4)$ *and* $\text{reach}(K) > 3r$. *For any two points* p *and* q *on* $\partial(K^r)$, *if* $\angle px_0q < \pi/10$, *then*

$$\|p - q\| \leq 7r \sin \angle px_0q.$$

Proof. Without loss of generality, we assume

$$\|p - x_0\| \geq \|q - x_0\|.$$

Because $q \in \partial(K^r)$, there exists a point $q_0 \in K$ such that $\|q - q_0\| = r$. Since $\text{reach}(K) > 3r$, by Lemma 32, the closed ball $B(w, 3r)$ centered at $w = q_0 + 3(q - q_0)$ with radius $3r$ intersects K only at q_0. Consequently, the closed ball $B(w, 2r)$ intersects K^r only at q. In particular, this means that p is outside the closed ball $B(w, 2r)$.

Since for any $u \in B(x_0, \frac{7}{10}r)$, we have

$$\|u - q_0\| \leq \|u - x_0\| + \|x_0 - q_0\| \leq \frac{7}{10}r + \frac{1}{4}r < r,$$

which implies that

$$\|w - u\| \geq \|w - q_0\| - \|u - q_0\| > 3r - r = 2r.$$

Thus, the ball $B(x_0, \frac{7}{10}r)$ does not intersect the ball $B(w, 2r)$. Also, because p is on the boundary of K^r, there exists $p_0 \in K$ such that $\|p - p_0\| = r$. Consequently,

$$\|p - x_0\| \geq \|p - p_0\| - \|p_0 - x_0\| \geq r - \frac{1}{4}r = \frac{3}{4}r > \frac{7}{10}r. \tag{3.12}$$

This means that p is outside the ball $B(x_0, \frac{7}{10}r)$. Therefore, we have two disjoint balls $B(x_0, \frac{7}{10}r)$ and $B(w, 2r)$, with the point q on the boundary of $B(w, 2r)$, and the point p lying outside both balls.

Let H be the two-dimensional plane containing x_0, p and q. Let Γ be the intersection of H with the boundaries of the balls $B(w, 2r)$, and γ be the intersection of H with the boundary of $B(x_0, \frac{7}{10}r)$. Clearly, γ is a circle centered at x_0 with radius $\frac{7}{10}r$. Since Γ contains q, Γ is not empty. Thus, Γ must be a circle (or the single point q in the extreme case, which can be viewed as a circle of radius 0). Let w' be the center of Γ.

Because $q_0 \in K \subset B(x_0, \frac{1}{4}r)$, the distance between q_0 and x_0 is at most $\frac{1}{4}r$. Thus, the distance between q_0 and the plane H is at most $\frac{1}{4}r$.

We claim that the distance from w to H is at most $\frac{r}{2}$. Note that w, q_0 and q are collinear with

$$\|w - q\| = 2\|q - q_0\|.$$

Let q_0' denote the closet point on H to q_0, and let w' denote the closest point on H to w. Thus, w' is the center of the circle Γ. (See figure 3.1.) Then, the triangle $\triangle q_0 q_0' q$ is similar to the triangle $\triangle w w' q$. In particular, this implies that

$$\frac{\|q_0 - q\|}{\|q_0 - q_0'\|} = \frac{\|w - q\|}{\|w - w'\|}. \tag{3.13}$$

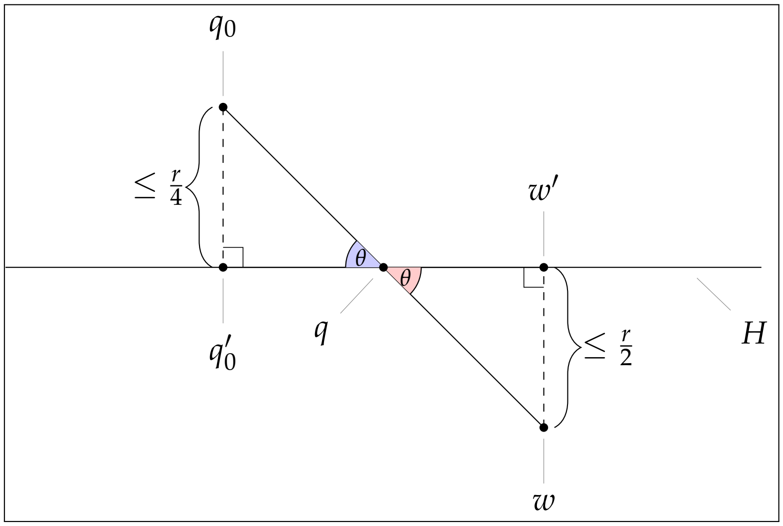

FIGURE 3.1: Lemma 36: Distance to H

Since, $\|w - q\| = 2\|q - q_0\|$, and $\|q_0 - q\| \leq \frac{1}{4}r$, we obtain

$$\|w - w'\| = 2\|q_0 - q\| \leq \frac{r}{2}.$$

We thus conclude that the distance between w and H is at most $\frac{1}{2}r$, i.e. $\|w - w'\| \leq \frac{1}{2}r$.

Next, we show that the radius of Γ is at least $\frac{\sqrt{15}}{2}r$. By considering the right triangle $\triangle ww'q$ we see that

$$\text{radius of } \Gamma = \sqrt{\|w - q\|^2 - \|w - w'\|^2} \tag{3.14}$$

$$\geq \sqrt{(2r)^2 - \left(\frac{r}{2}\right)^2} \tag{3.15}$$

$$= \frac{\sqrt{15}}{2}r. \tag{3.16}$$

Recall that p lies outside the ball $B(x_0, \frac{7}{10}r)$ and the ball $B(w, 2r)$, implying that p is outside the circles γ and Γ. Thus, there are two tangent lines from p to the circle γ and hence two tangent points, which we call t and s.

Consider the right triangle $\triangle x_0tp$. By equation (3.12) and the fact that$\|t - x_0\| = \frac{7}{10}r$, we have then have

$$\cos\angle tx_0p = \frac{\|t - x_0\|}{\|p - x_0\|} \leq \frac{\frac{7}{10}r}{\frac{3}{4}r} = \frac{14}{15}, \tag{3.17}$$

implying that

$$\angle tx_0p \geq \cos^{-1}\frac{14}{15} \geq \frac{\pi}{10}.$$

On the other hand, since

$$\|p - x_0\| \leq \|p - p_0\| + \|p_0 - x_0\| \leq r + \frac{1}{4}r = \frac{5}{4}r,$$

we have

$$\cos\angle tx_0p = \sin\angle tpx_0 = \frac{\|t - x_0\|}{\|p - x_0\|} \geq \frac{\frac{7}{10}r}{\frac{5}{4}r} = \frac{14}{25} > \frac{1}{2}, \tag{3.18}$$

which implies that $\angle tx_0p \leq \frac{\pi}{3}$. In conclusion, we have

$$\frac{\pi}{10} \leq \angle tx_0p \leq \frac{\pi}{3}.$$

Of course, we also have

$$\frac{\pi}{10} \leq \angle sx_0p \leq \frac{\pi}{3}.$$

Now, we estimate the angle $\angle qpx_0$. There are two cases. Case one: q and x_0 lies on different sides of one of the tangent lines, say line pt. Case two, q is between the rays pt and ps emanating from p.

In the first case, we have

$$\angle tpx_0 \leq \angle qpx_0.$$

Since we have assumed that

$$\|p - x_0\| \geq \|q - x_0\|,$$

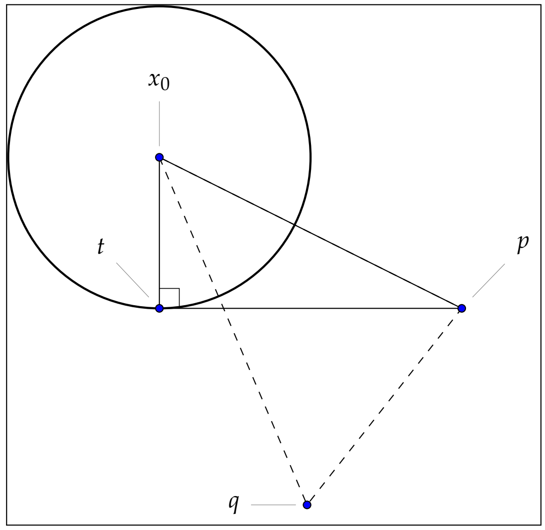

FIGURE 3.2: Lemma 36: Case 1

we also have $\angle qpx_0 < \pi/2$. Hence,

$$\frac{\pi}{10} \leq \angle qpx_0 \leq \frac{\pi}{2}. \tag{3.19}$$

Now, we consider the second case, that is, q lies in the cone between the rays pt and ps. Because by our initial hypothesis

$$\angle qx_0p < \pi/10 \leq \angle tx_0p,$$

q must lie in the region bounded by the line segment pt, line segment ps and the short arc st. This region has diameter less than $\frac{5}{4}r$, which is too small to contain the entire circle Γ. Together with the fact that p lies outside Γ, we conclude that Γ must intersect with one of the tangent lines twice. We assume that Γ intersects with pt at u and v, with v next to p. Let m be the midpoint of the line segment from u to v. Note that $\angle tpq \leq \angle uvq$. Since $\angle uvq$ equals half of the radian measure of the short arc uq on the circle Γ, which is less that half of the radian measure of the arc uqv on the circle, and the latter equals half of the angle $\angle uw'v$, or equals the angle $\angle mw'v$, we thus have

$$\angle tpq \leq \angle mw'v. \tag{3.20}$$

Note that

$$\begin{aligned} \|m - v\| &= \frac{1}{2} \|u - v\| && (3.21) \\ &\leq \frac{1}{2} \|t - p\| && (3.22) \\ &\leq \frac{1}{2} \|x_0 - p\| && (3.23) \\ &\leq \frac{5}{8} r. && (3.24) \end{aligned}$$

Also, recall from equation (3.16) that $\|w - v\| \geq \frac{\sqrt{15}}{2} r$. Thus, we have

$$\sin \angle m w' v = \frac{\|m - v\|}{\|w' - v\|} \leq \frac{\frac{5}{8} r}{\frac{\sqrt{15}}{2} r} \leq \frac{1}{3}. \tag{3.25}$$

Hence, we have

$$\sin \angle tpq \leq \frac{1}{3}. \tag{3.26}$$

Consequently,

$$\begin{aligned} \sin \angle qpx_0 &= \sin(\angle tpx_0 - \angle tpq) \\ &= \sin \angle tpx_0 \cos \angle tpq - \cos \angle tpx_0 \sin \angle tpq \\ &\geq \frac{1}{2} \cdot \sqrt{1 - \left(\frac{1}{3}\right)^2} - \sqrt{1 - \left(\frac{1}{2}\right)^2} \cdot \frac{1}{3} \\ &= \frac{2\sqrt{2} - \sqrt{3}}{6}. && (3.27) \end{aligned}$$

Therefore, in either case, we have

$$\sin \angle qpx_0 \geq \frac{2\sqrt{2} - \sqrt{3}}{6}.$$

By the law of sines applied to triangle $\triangle px_0q$, together with the fact that

$$\|x_0 - q\| \leq \frac{5}{4} r$$

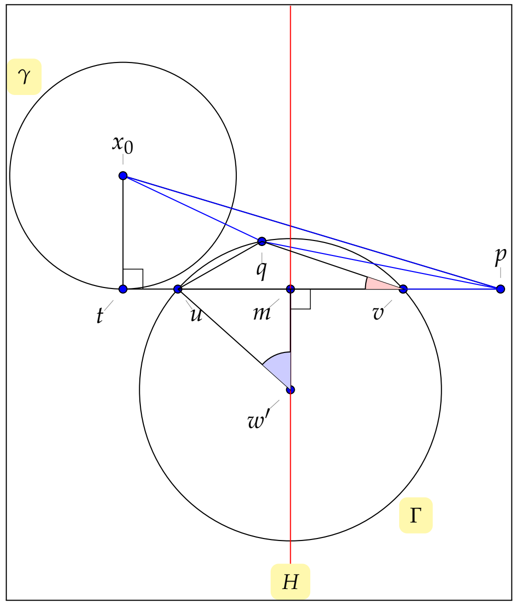

Figure 3.3: Lemma 36: Case 2

and the numerical calculation that

$$\frac{6}{2\sqrt{2}-\sqrt{3}} \cdot \frac{5}{4} < 7,$$

we have

$$\begin{aligned} \|p-q\| =& \frac{\|x_0-q\|}{\sin\angle qpx_0} \cdot \sin\angle px_0q \\ \leq& \frac{6}{2\sqrt{2}-\sqrt{3}}\frac{5}{4} r\sin\angle px_0q \\ \leq& 7r\sin\angle px_0q. \end{aligned} \tag{3.28}$$

□

3.5 SIMPLICIAL APPROXIMATION

Given a bounded closed star-shaped set L in $\mathbb{R}^d$ with vantage point x_0, we say a set S is a *δ-simplicial approximation of* L, if S is a star-shaped set with vantage point x_0, such that the boundary of S is a union of $(d-1)$-simplices with non-overlapping interiors. Furthermore, the vertices of each simplex are on ∂L, and the edge-length of each simplex is at most δ.

Lemma 37. *Suppose* $K \subset \mathbb{R}^d$ *with* $\mathrm{reach}(K) > c$. *For any* $0 < r \leq \frac{c}{4}$, *if* $\mathrm{diam}(K) \leq \frac{1}{2}r$, *then for every* $0 < \delta < r$, *and every* $\sqrt{r\delta}$*-simplicial approximation* S *of* K^r, *we have* $h(K^r, S) \leq \delta$.

Proof. By definition of Hausdorff distance, in order to establish the conclusion, we must show that δ satisfies

$$K^r \subseteq S + B(\delta), \tag{3.29}$$

and

$$S \subseteq K^r + B(\delta). \tag{3.30}$$

We begin by establishing equation (3.29). For any $y \in K^r \setminus S$, there exists $x_0 \in K$, so that $\|x_0 - y\| \leq r$. Because

$$\mathrm{diam}(K) \leq \frac{1}{2}r,$$

we have that

$$K \subset B(x_0, r/2).$$

The ray emanating from x_0 and passing through y intersects a facet of S at some point x. Let $x_1, x_2, \ldots, x_d$ be the vertices on that facet. Then x can be expressed as convex combination of x_i, $1 \leq i \leq d$. There are two cases: y lies between x_0 and x, or x lies between x_0 and y.

We first consider the former. Suppose y lies between x_0 and x. Since S is star-shaped with vantage point x_0, the entire line segment x_0x is contained in S. Hence, $y \in S$, and we immediately have $y \in S + B(\delta)$. On the other hand, suppose x lies between x_0 and y. Since the x_i are on the boundary of K^r, we must have

$$\|x_i - x_0\| \geq r.$$

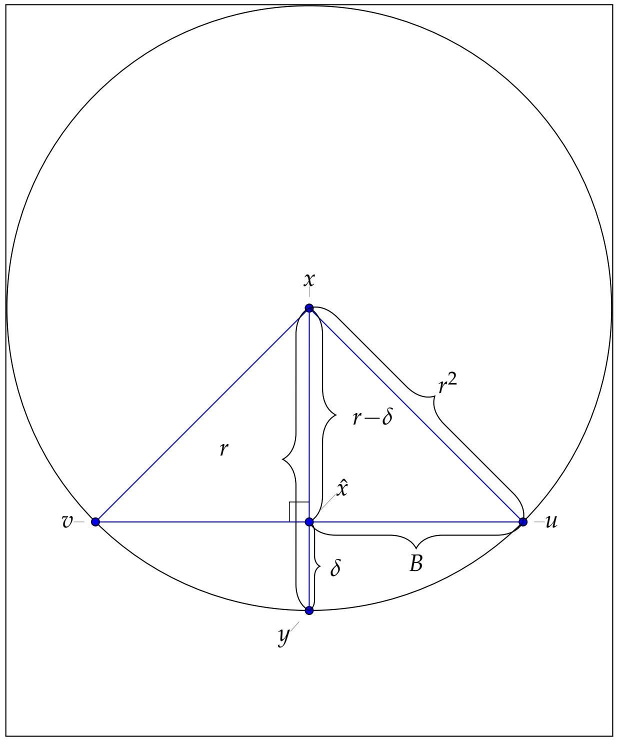

Figure 3.4: Spherical Cap Argument

Since

$$K \subseteq B(x_0, \frac{r}{2})$$

by hypothesis, Lemma 35 then implies that the entire line segment x_0x_i lies in K^r. Therefore, there exist points $\hat{x}_i$ on the line segment x_0x_i so that $\|\hat{x}_i - x_0\| = r$. For any $1 \leq i < j \leq d$, the triangle $\triangle x_i x_0 x_j$ contains the isosceles triangle $\triangle \hat{x}_i x_0 \hat{x}_j$, so we have

$$\|\hat{x}_i - \hat{x}_j\| \leq \|x_i - x_j\| \leq \sqrt{r\delta}.$$

Let $\triangle$ be the convex hull of $\hat{x}_i$, $1 \leq i \leq d$. The line segment x_0y cuts $\triangle$ at some point, say at $\hat{x}$. Since x_0 is a vantage point of K^r, Lemma 35 implies that $\hat{x} \in K^r$. We claim that

$$\|\hat{x} - x_0\| > r - \delta.$$

Indeed, suppose

$$\|\hat{x} - x_0\| \leq r - \delta.$$

Let L be the hyperplane passing through $\hat{x}$ and perpendicular to the line segment x_0y. Consider the spherical cap cut off by L from the ball centered at x_0 with radius r. Because the base radius of the spherical cap is at least

$$\sqrt{r^2 - (r-\delta)^2} = \sqrt{2r-\delta}\sqrt{\delta} \tag{3.31}$$

$$> \sqrt{r\delta}, \tag{3.32}$$

while

$$\|\hat{x}_i - \hat{x}\| \le \max_{j \ne i} \|\hat{x}_i - \hat{x}_j\| \tag{3.33}$$

$$\le \sqrt{r\delta}, \tag{3.34}$$

we see that the $\hat{x}_i$ all lie on the spherical cap but not on the base of the spherical cap. Consequently, as a convex combination of $\hat{x}_i$, $1 \le i \le d$, the point $\hat{x}$ does not lie on the base of the spherical cap. This is a contradiction. Hence

$$\|\hat{x} - x_0\| > r - \delta,$$

and consequently,

$$\|y - \hat{x}\| = \|y - x_0\| - \|\hat{x} - x_0\| \le \delta.$$

This implies that $y \in S + B(\delta)$.

Therefore, in either case, we have $y \in S + B(\delta)$, which implies that $K^r \subset S + B(\delta)$.

Next, we prove $S \subset K^r + B(\delta)$. For any x on the boundary of S, x belongs to a facet of S. Thus, there exist vertices x_i, $1 \le i \le d$ with

$$\|x_i - x_j\| \le \sqrt{r\delta}$$

for $1 \le i < j \le d$, such that x is a convex combination of x_i, $1 \le i \le d$. Thus

$$\operatorname{dist}(x, K) \le \|x - x_1\| + \operatorname{dist}(x_1, K) \le \sqrt{r\delta} + r < c. \tag{3.35}$$

Since $\text{reach}(K) > c$, there exists a unique point $x_0 \in K$ that is closest to x, and the closed ball $B(w, c)$ with center at w intersects K only at x_0, where we define

$$w := x_0 + \frac{c}{\|x - x_0\|}(x - x_0).$$

Consequently, the closed ball $B(w, c - r)$ with center at w and radius $c - r$ intersects K^r only one point. Denote this point by y. We have

$$\text{dist}(x, K^r) = \|x - y\|.$$

Clearly, y lies on the line segment from x_0 to w.

If x is between y and x_0, then, because K is star-shaped with respect to x_0, we have that the entire line segment from x_0 to y must lie in K^r, implying that $x \in K^r$. This immediately implies that $x \in K^r + B(\delta)$.

Suppose y is between x and x_0. We show that $\|x - y\| \le \delta$. Assume $\|x - y\| > \delta$. Consider the hyperplane H that passes through x and is perpendicular to the line segment xy. Since H contains x which is a linear combination of x_i, $1 \le i \le d$, there exists at least one x_i such that either $x_i \in H$ or x_i and y lie on different sides of H. Since

$$B(w, c - r) \bigcap K^r = \{y\},$$

this x_i lies outside the ball $B(w, c - r)$. Thus,

$$\begin{aligned}
\|x_i - x\| \ge& \sqrt{(c - r)^2 - (c - r - \|x - y\|)^2} && (3.36)\\
\ge& \sqrt{2c - 2r - 2\delta}\sqrt{\delta} && (3.37)\\
>& \sqrt{r\delta}. && (3.38)
\end{aligned}$$

However, since

$$\|x - x_i\| \le \max_{j \ne i} \|x_i - x_j\| \le \sqrt{r\delta},$$

we get a contradiction. Therefore, we must have $\|x - y\| \le \delta$, which implies that $x \in K^r + B(\delta)$. Hence, in either case, we have $x \in K^r + B(\delta)$. Because x is an arbitrary point on ∂S, we have $\partial S \subset K^r + B(\delta)$. Note that S and $K^r + B(\delta)$ are star-shaped with

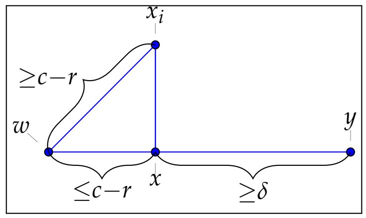

Figure 3.5: Equations (3.36) Through (3.38)

common vantage point x_0. We conclude that

$$S \subset K^r + B(\delta).$$

□

3.6 local approximation

In this section, we approximate sets that are already close to a fixed set.

Lemma 38. *Let $\mathcal{K}$ be the collection of all sets contained in $B(1)$ with $\text{reach}(K) > c$. For each fixed $D \in \mathcal{K}$, and $0 < \eta < \frac{3}{4}\sin(\frac{\pi}{10})r$, where $r = \frac{c}{3}$, denote*

$$\mathcal{D}_\eta = \{C \in \mathcal{K} \mid h(C, D) < \eta\}.$$

Then,

$$\log N\left(\frac{\eta}{2}, \mathcal{K}, h\right) \leq \lambda(c, d)\eta^{-(d-1)/2}$$

where $\lambda(c, d)$ is a constant depending only on c and d.

Proof. Let M be the maximum number of points that can be choosen from $B(1)$ with mutual distance at least $r/4$. Then, the balls centered at these points with radius $r/4$ cover $B(1)$, while the balls centered at these points with radius $r/8$ are disjoint and contained in $B(1 + r/8)$. The total volume of the M small balls is

$$M \cdot (r/8)^d \cdot \text{vol}(B(1)),$$

while the volume of $B(1 + r/8)$ is

$$(1 + r/8)^d \cdot \mathrm{vol}(B(1).$$

By comparing the volumes, we obtain

$$M \leq \left(1 + \frac{8}{r}\right)^d .$$

This means that $B(1)$ can be covered by no more than $(1 + 24/c)^d$ closed balls of radius $r/4$. Denote these balls by $B(c_i, r/4)$, $1 \leq i \leq M$. By Proposition 30, for each $K \in \mathcal{K}$, $\mathrm{reach}(K \cap B(c_i, r/4)) > c$. Consider the collection

$$\mathcal{K}_i(\eta) := \left\{ K \cap B(c_i, r/4) \mid K \in \mathcal{D}_\eta \right\} .$$

If we denote $D_i = D \cap B(c_i, r/4)$, then

$$\begin{aligned} \mathcal{K}_i(\eta) &\subset \{K \subset B(c_i, r/4) \mid \mathrm{reach}(K) > c, h(K, D_i) \leq \eta\} \\ &= \left\{ c_i + K \mid K \subset B(0, r/4), \mathrm{reach}(K) > c, h(K, D_i') \leq \eta \right\} , \end{aligned}$$

where $D_i' = (D_i - c_i) \cap B(r/4)$. We denote

$$\mathcal{E}_\eta^i = \{K \subset B(x_i, r/4) \mid \mathrm{reach}(K) > c, h(K, D_i) \leq \eta\} .$$

Now, for each $K \in \mathcal{E}_\eta^i$, we approximate K^r with a simplicial sphere. Note that

$$K^r \subset B(r/4) + B(r) \subset \left[-\frac{5}{4} r, \frac{5}{4} r \right]^d .$$

The $2d$ facets of $[-\frac{5}{4}r, \frac{5}{4}r]^d$ can be triangulated into $\mu\eta^{-(d-1)/2}$ simplices of dimension $(d-1)$ and edge length at most $\frac{1}{24}\sqrt{r\eta}$, where μ is a constant depending only on d and c. Let v_i, $1 \leq i \leq d$ be the vertices of these simplices. Thus,

$$\|v_i - 0\| \geq \frac{5}{4} r$$

for all $1 \leq i \leq d$. Since we have proved that K^r must be star-convex with vantage point 0 in Lemma 35, the line segment connecting each v_i and the origin intersects the boundary of K^r at a unique point, say s_i. Let S_K be the simplicial sphere with vertices s_i, $1 \leq i \leq m$. We show that S is a $\sqrt{r\eta}$-simplicial approximation of K^r. Indeed, for any two adjacent vertices v_i and v_j, because

$$\|v_i - v_j\| \leq \frac{1}{24}\sqrt{r\eta},$$

and $\|v_i\| \geq \frac{5}{4}r$, $\|v_j\| \geq \frac{5}{4}r$, we have

$$\sin(\theta/2) \leq \frac{1}{30}\sqrt{r\eta},$$

where θ is the angle between the vector v_i and v_j. This implies that

$$\sin\theta \leq \frac{1}{15}\sqrt{r\eta}.$$

By Lemma 36, we have

$$\left\|s_i - s_j\right\| \leq \frac{1}{2}\sqrt{r\eta}.$$

Applying Lemma 37, we obtain

$$h(K, S_K) \leq \frac{1}{4}\eta.$$

The simplicial sphere S_K we obtained above cannot be directly used to construct $\eta/2$-net for $\mathcal{E}^i_\eta$ because the choices of the vertices depends on K too specifically. In what follows, we will reduce its dependence on K.

Let t_i be the intersection of the line segment $0v_i$ with the boundary of D^r_i. Then s_i, t_i, v_i and the origin are on the same line. We claim that $\|t_i - s_i\| \leq 11\eta$.

Without loss of generality, we assume s_i lies between t_i and 0. Because

$$h(K^r, D^r_i) = h(K, D_i) \leq \eta,$$

there exists $x_i \in K^r$ such that
$$\|t_i - x_i\| \le \eta.$$

Now, let us look at the $\triangle t_i 0 x_i$. We have

$$\|t_i - 0\| \ge \text{dist}(t_i, D_i) - \text{dist}(0, D_i) \ge r - \frac{1}{4}r = \frac{3}{4}r,$$

$$\|x_i - 0\| \ge \text{dist}(x_i, K) - \text{dist}(0, K) \ge r - \frac{1}{4}r = \frac{3}{4}r,$$

and
$$\|t_i - x_i\| \le \eta \le \frac{3}{4}\sin(\frac{\pi}{10})r.$$

Therefore, a simple computation gives

$$\sin \angle t_i 0 x_i \le \frac{4\eta}{3r} \le \sin(\frac{\pi}{10}). \tag{3.39}$$

Hence,
$$\angle t_i 0 x_i \le \frac{\pi}{10}.$$

Now, we consider the triangle $\triangle s_i 0 x_i$. Because

$$\|s_i - 0\| \ge \text{dist}(s_i, K) - \text{dist}(0, K) \ge r - \frac{1}{4}r = \frac{3}{4}r,$$

$\|x_i - 0\| \ge \frac{3}{4}r$, and
$$\angle s_i 0 x_i = \angle t_i 0 x_i \le \frac{\pi}{10},$$

we can apply Lemma 36, and use (3.39) to obtain

$$\|s_i - x_i\| \le 7r \sin \angle s_i 0 x_i \le 10\eta.$$

Therefore,
$$\|t_i - s_i\| \le \|t_i - x_i\| + \|x_i - s_i\| \le 11\eta.$$

This finishes the proof of the claim.

Now, for an integer $k \leq 44$, we choose $\hat{s}_i$ to be the point on the line segment $0v_i$ that is closest to s_i and with distance $k\eta/4$ way from t_i Then it is clear, there are less than 100 choices for each $\hat{s}_i$. Let $\hat{S}_K$ be the simplicial sphere with vertices $\hat{s}_i$. Clearly, we have

$$h(S_i, \hat{S}_i) \leq \frac{\eta}{4},$$

which implies that

$$h(K, \hat{S}_K) \leq \frac{1}{2}\eta.$$

Since there are less than 100 choices for each $\hat{s}_i$, the total number of choices of $\widehat{S}_K$ is no more than

$$100^{\mu\eta^{-(d-1)/2}} \leq \exp\left(C\eta^{-(d-1)/2}\right),$$

where C is a constant depending only on d and c. This means that

$$N(\eta/2, \mathcal{E}_\eta^i, h) \leq \exp\left(C\eta^{-(d-1)/2}\right).$$

In particular, this implies that for each $1 \leq i \leq M$,

$$N(\eta/2, \mathcal{K}_i(\eta), h) \leq \exp\left(C\eta^{-(d-1)/2}\right).$$

Now we turn to estimate $N(\eta/2, \mathcal{D}_\eta, h)$. For each $1 \leq i \leq M$, there exists an $\eta/2$-net of $\mathcal{K}_i(\eta)$ of cardinality $\exp\left(C\eta^{-(d-1)/2}\right)$. Denote these nets by $\mathcal{N}_i$.

For each $K \in \mathcal{D}_\eta$, we have $K \cap B(c_i, r/4) \in \mathcal{K}_i(\eta)$. Thus, there exists $S_i \in \mathcal{N}_i$ such that

$$h(K \cap B(c_i, r/4), S_i) \leq \frac{\eta}{2}.$$

Now we define

$$S = \cup_{i=1}^M S_i.$$

This S may not have positive reach. So, we define $\widetilde{S}$ to be intersection of all sets $P \supset S$ with $\mathrm{reach}(P) > 2r$. Since $\mathrm{reach}(K) > 3r$, and $\eta < r$, we have

$$\mathrm{reach}(K^{\eta/2}) \geq \mathrm{reach}(K) - \frac{\eta}{2} > 2r,$$

and $K^{\eta/2} \supset S_i$ for all $1 \le i \le M$, we have

$$P \subset K^{\eta/2} = K + B(\eta/2).$$

On the other hand, we clearly have

$$K \subset \cup_{i=1}^{M} S_i + B(\eta/2) \subset P + B(\eta/2).$$

Thus,

$$h(K, P) \le \eta/2.$$

Now the total number of choices of P is no more than

$$\exp\left(C\eta^{-(d-1)/2}\right)^{M} \le \exp\left(\lambda(c,d)\eta^{-(d-1)/2}\right),$$

where $\lambda(c,d)$ is a constant depending only on c and d.

This proves Lemma 38. □

3.7 PROOF OF THEOREM 33

Proof. Let $P(n)$ denote the statement:

$$\log N(\varepsilon, \mathcal{K}, h) \le \Lambda(c,d) e^{-\frac{d-1}{2}}, \tag{3.40}$$

for all $\varepsilon \ge 2^{-n}$, with $n \in \mathbb{N} \bigcup \{0\}$. By choosing a sufficiently large constant, we only need to prove for the case when $\varepsilon \le \frac{3}{4}\sin(\pi/10)r$ so we can directly use the previous lemma. We prove Theorem 33 by induction on $n \ge n_0 := \lceil -\log_2(\frac{3}{4}\sin(\pi/10)r)\rceil$.

Again, we may choose a sufficiently large constant, say $\Lambda(d,c) > 2\lambda(d,c)$, to ensure that the statement $P(n_0)$ is true. Now assume that the statement $P(n)$ is true for all $\varepsilon \ge 2^{-n}$, for some fixed $n \in \mathbb{N}$. We will show that $P(n+1)$ also holds. By the induction hypothesis, $P(n)$, there exists an ε-net of $\mathcal{K}$, say, $D_1, \ldots, D_N$, with

$N := N(\varepsilon, \mathcal{K}, h)$. We approximate each of these D_i under the Haüsdorff norm. Let

$$\mathcal{K}_i := \{C \in \mathcal{K} \mid h(C, D_i) \leq \varepsilon\}.$$

Clearly $\mathcal{K} \subset \bigcup_{n=1}^{N} \mathcal{K}_i$. By Lemma 38,

$$N(\varepsilon/2, \mathcal{K}_i, h) \leq \exp\left(\lambda(c,d)\varepsilon^{-(d-1)/2}\right).$$

Then,

$$N(\frac{\varepsilon}{2}, \mathcal{K}, h) \leq N(\frac{\varepsilon}{2}, \mathcal{K}_i, h) \cdot N \tag{3.41}$$

$$= N(\frac{\varepsilon}{2}, \mathcal{K}_i, h) \cdot \exp\left(\mu(c,d)\left(\frac{\varepsilon}{2}\right)^{-(d-1)/2}\right) \tag{3.42}$$

$$= \exp\left(\lambda(c,d)\varepsilon^{-(d-1)/2}\right)\exp\left(\mu(c,d)\left(\frac{\varepsilon}{2}\right)^{-(d-1)/2}\right) \tag{3.43}$$

$$= \exp\left(\Lambda(c,d)\left(\frac{\varepsilon}{2}\right)^{-(d-1)/2}\right), \tag{3.44}$$

where $\mu(c,d)$ is the constant guranteed by the induction hypothesis $P(n)$ and

$$\Lambda(c,d) := \lambda(c,d) + \mu(c,d).$$

Thus the statement holds for all $\varepsilon > 0$. □

3.8 EXTENSION

The following theorem is readily obtained from the proof of Theorem 33.

Theorem 39. *Let $\mathcal{P}_m$ be the collection of d-dimensional sets contained in $B(R)$ that can be expressed as the union of at most m sets of reach greater than c. The there exists a constant $\Lambda(c,d)$ depending only on c and d, such that for every* $0 < \varepsilon < 1$,

$$\log N(\varepsilon, \mathcal{P}_m, h) \leq \Lambda(c,d) m R^d \varepsilon^{-(d-1)/2}.$$

CHAPTER 4

Metric Entropy of Q-hulls in Hilbert Spaces

In this chapter, we consider the following problem: Given a pre-compact subset T of a Hilbert space H, suppose we have known the metric entropy of the set T. What can we say about the metric entropy of its q-hull, $0 < q \leq 1$, defined by

$$H_q(T) := \left\{ \sum_{i=1}^{n} \alpha_i t_i \;\middle|\; t_i \in T, 1 \leq i \leq n, n \geq 1, \sum_{i=1}^{n} |\alpha_i|^q \leq 1 \right\}? \tag{4.1}$$

4.1 A PRELIMINARY RESULT

We first quote a result in special case of Proposition 13 of Chapter 2 that will be used in the later proof:

Proposition 40. *([9], p. 98) Fix $n \in \mathbb{N}$ and let ℓ_q^n and ℓ_1^n denote $\mathbb{R}^n$ under the ℓ_q and ℓ_1 metrics respectively. Let $id\colon \ell_q^n \to \ell_1^n$ be the identity map, we abbreviate $e_k = e_k(id)$. Then, for $0 < q \leq 1$,*

$$e_k \leq c \cdot \begin{cases} 1, & q \leq k \leq \log(2n), \\ \left[k^{-1} \log\left(1 + \frac{2n}{k}\right)\right]^{\frac{1}{q}-1}, & \log(2n) \leq k \leq 2n, \\ 2^{\frac{-k}{2n}} (2n)^{\frac{1}{q}-1}, & k \geq 2n, \end{cases} \cdot$$

Proposition 40 implies the following proposition stated in terms of metric entropy.

Proposition 41. *Let $B_{\ell_q^n}$ be the unit ball of ℓ_q^n, $0 < q \leq 1$. Then the ε-metric entropy of $B_{\ell_q^n}$, under the ℓ_1 norm, is given by*

$$\log N(\varepsilon, B_{\ell_q^n}, \|\cdot\|_{\ell_1^n}) \leq \begin{cases} C\varepsilon^{-q/(1-q)} \log(2n\varepsilon^{q/(1-q)}), & n\varepsilon^{q/(1-q)} > 1 \\ \\ Cn \log(2n\varepsilon^{-q/(1-q)}), & n\varepsilon^{q/(1-q)} < 1 \end{cases}. \tag{4.2}$$

Proof. If $\varepsilon \geq 1$, then the left-hand side of inequality (4.2) is 0, so inequality (4.2) is clearly true. We thus consider the case $\varepsilon < 1$. If $n\varepsilon^{q/(1-q)} > 1$, we choose k to be the smallest integer such that

$$\varepsilon \leq c \left[k^{-1} \log\left(1 + \frac{2n}{k}\right)\right]^{\frac{1}{q}-1}. \tag{4.3}$$

where c is the constant in Proposition 40. Then, we choose k to be the smallest integer such that

$$\begin{aligned} n\varepsilon^{q/(1-q)} \geq & c^{q/(1-q)} \frac{n}{(k+1)} \log\left(1 + \frac{2n}{(k+1)}\right) \\ \geq & \frac{n}{c_1 k} \log\left(1 + \frac{2n}{c_1 k}\right), \end{aligned} \tag{4.4}$$

for some integer c_1. Since we can choose c_1 so that

$$\log(2n) \leq c_1 k \leq 2n,$$

by Proposition 40, we have

$$N(\varepsilon, B_{\ell_q^n}, \|\cdot\|_{\ell_1^n}) \leq 2^{c_1 k - 1}. \tag{4.5}$$

Because inequality (4.3) implies that

$$\varepsilon^{\frac{q}{1-q}} \asymp \frac{1}{k} \log\left(1 + \frac{2n}{k}\right) \asymp \frac{1}{k} \log\left(\frac{2n}{k}\right). \tag{4.6}$$

Thus,

$$\log(2n\varepsilon^{\frac{q}{1-q}}) \asymp \log\left(\frac{2n}{k}\right) + \log\log\left(\frac{2n}{k}\right) \asymp \log\left(\frac{2n}{k}\right). \tag{4.7}$$

Plugging into (4.6), we have

$$k \asymp \varepsilon^{-q/(1-q)} \log(2n\varepsilon^{\frac{q}{q-1}}). \tag{4.8}$$

Plugging into (4.16), we have

$$\log N(\varepsilon, B_{\ell_q^n}, \|\cdot\|_{\ell_1^n}) \leq C\varepsilon^{-q/(1-q)} \log(2n\varepsilon^{\frac{q}{q-1}}).$$

Now we look at the the case $n\varepsilon^{q/(1-q)} \leq 1$. We can choose k, $k \geq 2n$, to be the smallest integer such that

$$\varepsilon \geq c2^{-\frac{k}{2n}} (2n)^{\frac{1}{q}-1}, \tag{4.9}$$

where c is the constant in Proposition 40. We can choose a positive integer c_2 such that

$$c2^{-\frac{k}{2n}} (2n)^{\frac{1}{q}-1} \geq 2^{-\frac{c_2 k}{2n}} (2n)^{\frac{1}{q}-1}.$$

Since

$$k \asymp 2n \log \frac{(2n)^{1-\frac{1}{q}}}{\varepsilon}, \tag{4.10}$$

we have

$$\log N\left(\varepsilon, B_{\ell_p^n}, \ell_q\right) \leq c_2 k \leq Cn \log(2n\varepsilon^{-q/(1-q)}).$$

This finishes the proof of the proposition. □

4.2 METRIC ENTROPY OF q-HULLS OF A FINITE SET

We first consider the case when $T = \{t_1, t_2, \ldots, t_n\}$ with $\|t_i\| \leq 1$ for all $1 \leq i \leq n$. We estimate $N(\varepsilon, H_q(T), \|\cdot\|_2)$, where

$$H_q(T) = \left\{ \sum_{i=1}^{n} c_i t_i \;\middle|\; \sum_{i=1}^{n} |c_i|^q \leq 1 \right\}. \tag{4.11}$$

Let

$$N_0 := N\left(\delta, B_{\ell_q}^n(T), \|\cdot\|_{\ell_2^n}\right)$$

So, there is a set of points $\{y_i\} \subset \mathbb{R}^n$ such that,

$$B_{\ell_q^n} \subset \bigcup_{i=1}^{N_0} B_{\ell_2^n}(y_i, \delta)$$

where

$$B_{\ell_2^n}(y, \delta) = \left\{ x \in \mathbb{R}^n \;\middle|\; \|x - y\|_{\ell_2^n} \leq \delta \right\}.$$

Thus, we have

$$H_q(T) = \left\{ \sum_{i=1}^{n} c_i t_i \;\middle|\; (c_1, \ldots, c_n) \in B_{\ell_q^n} \right\} \tag{4.12}$$

$$\subset \left\{ \sum_{i=1}^{n} c_i t_i \;\middle|\; (c_1, \ldots, c_n) \in \cup_{j=1}^{N_0} B_{\ell_2^n}(y_j, \delta) \right\} \tag{4.13}$$

$$= \bigcup_{j=1}^{N_0} \left(\sum_{i=1}^{n} y_j(i) t_i + \left\{ \sum_{i=1}^{n} c_i t_i \;\middle|\; \sum_{i=1}^{n} |c_i| \leq \delta \right\} \right) \tag{4.14}$$

$$= \bigcup_{j=1}^{N_0} \left(\sum_{i=1}^{n} y_j(i) t_i + \left\{ \delta \sum_{i=1}^{n} c_i t_i \;\middle|\; \sum_{i=1}^{n} |c_i| \leq 1 \right\} \right). \tag{4.15}$$

Note that

$$\left\{ \delta \sum_{i=1}^{n} c_i t_i \;\middle|\; \sum_{i=1}^{n} |c_i| \leq 1 \right\} = \delta \cdot H_1(T).$$

If $\mathcal{N}_0 = \{x_1, \ldots x_m\}$ is an η-net of $H_1(T)$. Then for each $j = 1, \ldots, N_0$, the set

$$\sum_{i=1}^{n} y_j(i) t_i + \delta \cdot \mathcal{N}_0 = \left\{ \sum_{i=1}^{n} y_j(i) t_i + \delta x_j \;\middle|\; j = 1, \ldots, m \right\}$$

is an $\delta\eta$-net of

$$\sum_{i=1}^{n} y_j(i) t_i + \left\{ \delta \sum_{i=1}^{n} c_i t_i \;\middle|\; \sum_{i=1}^{n} |c_i| \leq 1 \right\}.$$

Consequently, the set

$$\mathcal{N} := \cup_{j=1}^{N_0} \left(\sum_{i=1}^{n} y_j(i) t_i + \delta \cdot \mathcal{N}_0 \right)$$

is a $\delta\eta$-net of $H_q(T)$ with cardinality $N_0 m$. This implies that

$$N(\delta\eta, H_q(T), \|\cdot\|_2) \leq N_0 m. \tag{4.16}$$

Now, we turn to estimate m. We use a well-known probabilistic method due to Maurey. Define a random vector X by

$$\begin{aligned} 1) \qquad & \mathbb{P}\left(X = \text{sign}(c_i) \cdot t_i\right) := |c_i| \\ 2) \qquad & \mathbb{P}\left(X = 0\right) := 1 - \sum_{i=1}^{n} |c_i| \end{aligned}$$

Let $v = \sum c_i X_i$. Note that $\mathbb{E}\left(X\right) = v$, which is in $H_1(T)$. Let $X_1, \ldots, X_k; \widetilde{X}_1, \ldots, \widetilde{X}_k$ be i.i.d. copies of X. Define random vector Y by

$$Y = \frac{1}{k} \sum_{i=1}^{k} X_i.$$

Note that

$$\mathbb{E}\left(Y\right) = v = \mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} X_i\right) = \mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} \widetilde{X}_i\right).$$

Consider the following computations:

$$\begin{aligned} \mathbb{E}_\omega \|Y - v\|_2 = & \mathbb{E}_\omega \left\| Y - \mathbb{E}_{\widetilde{\omega}}\left(\frac{1}{k}\sum_{j=1}^{k} X_j\right)\right\|_2 \\ = & \mathbb{E}_\omega \left\| \mathbb{E}_{\widetilde{\omega}}\left(\frac{1}{k}\sum_{j=1}^{k} \widetilde{X}_j - \frac{1}{k}\sum_{j=1}^{k} X_j\right)\right\|_2 \\ \leq & \mathbb{E}_\omega \mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k}\sum_{j=1}^{k} \widetilde{X}_j - \frac{1}{k}\sum_{j=1}^{k} X_j \right\|_2 \\ = & \mathbb{E}_\omega \mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k}\sum_{j=1}^{k} (\widetilde{X}_j - X_j) \right\|_2 . \end{aligned} \tag{4.17}$$

Because $\widetilde{X}_j - X_j$ is symmetric in the product space $\widetilde{\Omega} \times \Omega$, if we let $\varepsilon_1, \ldots, \varepsilon_k$ be i.i.d. Bernoulli random variables, independent of X_j and $\widetilde{X}_j$, $1 \leq j \leq k$, with $P(\varepsilon_i = 1) =$

$P(\varepsilon_i = -1) = \frac{1}{2}$, then we have

$$\begin{aligned}
&\mathbb{E}_\omega \mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k} \sum_{j=1}^{k} (\widetilde{X}_j - X_j) \right\|_2 \\
=&\mathbb{E}_\omega \mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j (\widetilde{X}_j - X_j) \right\|_2 \\
\leq&\mathbb{E}_\omega \mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j \right\|_2 + \mathbb{E}_\omega \mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j) \right\|_2 \\
=&2\mathbb{E}_\omega \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j) \right\|_2 .
\end{aligned}$$

Plugging into (4.17), and taking expectation over the space of ε_i, we obtain

$$\begin{aligned}
\mathbb{E}_\omega \|Y - v\|_2 \leq& 2\mathbb{E}_\varepsilon \mathbb{E}_\omega \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j \right\|_2 && (4.18) \\
=&2\mathbb{E}_\omega \mathbb{E}_\varepsilon \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j \right\|_2 && (4.19) \\
\leq&2\mathbb{E}_\omega \left(\mathbb{E}_\varepsilon \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j \right\|_2^2 \right)^{1/2} . && (4.20)
\end{aligned}$$

Because

$$\mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j \right\|_2^2 = \sum_{j=1}^{k} \|X_j\|^2,$$

and $X_i \in \{\pm t_1, \dots, \pm t_n\}$ with $\|t_i\| \leq 1$ for all $1 \leq i \leq n$, we have

$$\mathbb{E}_{\widetilde{\omega}} \left\| \frac{1}{k} \sum_{j=1}^{k} \varepsilon_j X_j \right\|_2^2 \leq k.$$

This implies that

$$\mathbb{E}_\omega \|Y - v\|_2 \leq \frac{2}{\sqrt{k}}.$$

Thus, for any $v \in H_1(T)$, we have shown that u lies within a distance of $\frac{2}{\sqrt{k}}$ of a realization of Y. Note that there are $(2n+1)^k$ possible realizations of the Y. If we let

$\eta = \frac{2}{\sqrt{k}}$, then $k = \frac{4}{\eta^2}$, and we have

$$N\left(\eta, H_1(T), \|\cdot\|_2\right) \leq (2n+1)^k = (2n+1)^{4\eta^{-2}}.$$

Therefore, we can choose m as small as $(2n+1)^{4\eta^{-2}}$. Plugging this into (4.16) we obtain

$$N(\delta\eta, H_q(T), \|\cdot\|_2) \leq N_0 \cdot (2n+1)^{4\eta^{-2}}.$$

On the other hand, by Proposition 41, we can choose N_0 as small as

$$N_0 \leq (2n+1)^{C\delta^{-q/(1-q)}},$$

for some constant $C > 0$. Thus,

$$N(\delta\eta, H_q(T), \|\cdot\|_2) \leq (2n+1)^{C\delta^{-q/(1-q)}+4\eta^{-2}}.$$

In particular, if we choose $\delta = \eta^{2(1-q)/q}$. Then we obtain

$$N(\eta^{(2-q)/q}, H_q(T), \|\cdot\|_2) \leq (2n+1)^{(C+4)\eta^{-2}},$$

which implies that

$$\log N(\varepsilon, H_q(T), \|\cdot\|_2) \leq C\varepsilon^{-\frac{2q}{2-q}\log n}. \tag{4.21}$$

4.3 METRIC ENTROPY OF $H_q(T)$ WHEN T IS INFINITE

Now we consider the case when T is a infinite set. We consider the case when

$$\log N(\varepsilon, T, \|\cdot\|_2) \leq C\varepsilon^{-a} \tag{4.22}$$

for all $0 < \varepsilon < 1$, where $a > \frac{2q}{2-q}$ is some constant.

For any $0 < \varepsilon < 1$, and any integer $0 \le k \le m$, where m is the smallest integer such that $2^{m-1}\varepsilon > 1$. Let $\mathcal{N}_k$ be the $2^{k-1}\varepsilon$-net of T with minimum cardinality. Denote the cardinality of $\mathcal{N}_k$ by N_k. Thus, by our hypothesis (4.22)

$$N_k \le e^{C(2^{k-1}\varepsilon)^{-a}},$$

for all $0 \le k < m$. Furthermore $\mathcal{N}_m$ is only a singleton, say $\{t_m^*\}$. Note that $H_q(\mathcal{N}_0)$ is an $\varepsilon/2$-net of $H_q(T)$. If we can find $\varepsilon/2$-net of $H_q(\mathcal{N}_0)$ with cardinality N, then this $\varepsilon/2$-net of $H_q(\mathcal{N}_0)$ is an ε-net of $H_q(T)$. Therefore, it suffices to study

$$N(\varepsilon/2, H_q(\mathcal{N}_0), \|\cdot\|_2).$$

Because for each $t_0 \in \mathbb{N}_0$, we can write

$$t_0 = t_m + \sum_{i=1}^{m}(t_{i-1} - t_i),$$

where for all $1 \le i \le m$ $t_i \in \mathcal{N}_i$ is a point satisfying $\|t_i - t_{i-1}\| \le 2^{i-1}\varepsilon$, and $t_m = t_m^*$ is independent of t_0.

If we let

$$K_i = \left\{ t - s \;\middle|\; t \in \mathcal{N}_i, s \in \mathcal{N}_{i-1}, \|t-s\| \le 2^{i-2}\varepsilon \right\}.$$

Then each K_i has no more than $e^{2C(2^{i-2}\varepsilon)^{-a}}$ elements, and the norm of each elements is bounded by $2^{i-1}\varepsilon$. By what we have proved in the finite case, we have

$$\log N(\eta_i, H_q(K_i), \|\cdot\|_2) \le C\left(\frac{\eta_i}{2^{i-1}\varepsilon}\right)^{-2q/(2-q)} \cdot 2C(2^{i-2}\varepsilon)^{-a}. \tag{4.23}$$

Because,

$$H_q(\mathcal{N}_0) \subset t_m^* + \sum_{i=1}^{m} H_q(K_i),$$

if we choose η_i so that $\sum_{i=1}^{m} \eta_i = \varepsilon/2$, then we have

$$N(\varepsilon/2, H_q(\mathcal{N}_0), \|\cdot\|) \le \prod_{i=1}^{m} \log N(\eta_i, H_q(K_i), \|\cdot\|_2).$$

This implies that

$$\begin{aligned}\log N(\varepsilon, H_q(T), \|\cdot\|_2) \leq& \log N(\varepsilon/2, H_q(\mathcal{N}_0), \|\cdot\|_2)\\ \leq& \sum_{i=1}^{m}\left(C\left(\frac{\eta_i}{2^{i-1}\varepsilon}\right)^{-2q/(2-q)}\cdot 2C(2^{i-2}\varepsilon)^{-a}\right),\end{aligned}$$

where η_i are any positive numbers satisfying $\sum_{i=1}^{m}\eta_i = \varepsilon/2$. In particular, if we choose

$$\eta_i = \frac{i^{-2}\varepsilon}{2\sum_{i=1}^{m} i^{-2}} \geq \frac{3\varepsilon}{\pi^2 i^2},$$

then we have

$$\begin{aligned}\log N(\varepsilon, H_q(T), \|\cdot\|_2) \leq& \sum_{i=1}^{m}\left(C\left(\frac{3}{2^{i-1}\pi^2 i^2}\right)^{-2q/(2-q)}\cdot 2C(2^{i-2}\varepsilon)^{-a}\right)\\ \leq& 2^a C^2 \sum_{i=1}^{m}(2^{i-1})^{a-\frac{2q}{2-q}} i^{\frac{4q}{2-q}}\cdot\varepsilon^{-a}\\ \leq& C'\varepsilon^{-a},\end{aligned}$$

where C is a constant depending on $a - \frac{2q}{2-q}$.

In conclusion, we have proven the main result of this chapter:

Theorem 42. *If K is a precompact set in a Hilbert space satisfying*

$$\log N(\delta, K, \|\cdot\|) = O(\delta^{-\alpha}),$$

where for $\alpha > \frac{2q}{2-q}$, and $0 < q \leq 1$, then

$$\log N(\varepsilon, H_q(K), \|\cdot\|) = O\left(\varepsilon^{-\alpha}\right).$$

CHAPTER 5

Separately Con Functions

5.1 BACKGROUND

Shape-constrained functions are very important in applications, such as in nonparametric estimation of densities in statistics where many interesting classes of densities are shape-constrained. For a single variable function, monotonicity and convexity are the two most commonly used ones in practice. In general, one can study the so-called k-monotonic functions. Recall that a function on an interval $I \subset \mathbb{R}$ is called a k-monotonic function, if $(-1)^i f^{(i)}(x) \geq 0$ for all $x \in I$, and $0 \leq i \leq k-2$, and furthermore, $(-1)^{k-2} f^{(k-2)}$ is convex and $(-1)^{k-1} f^{(k-1)}$ is decreasing on I.

It is known (c.f. [32]) that if $\mathcal{M}$ is the class of monotonic functions on $[0,1]$ that are bounded by 1, then for any $1 \leq p < \infty$,

$$\log N_{[\,]}(\varepsilon, \mathcal{M}, \|\cdot\|_p) \asymp \log N(\varepsilon, \mathcal{M}, \|\cdot\|_p) \asymp \varepsilon^{-1}.$$

It is also proved (c.f. [13] and [17]) that if $\mathcal{M}^k$ denotes the class of k-monotonic on $[0,1]$ that are bounded by 1, than for any $1 \leq p < \infty$,

$$\log N_{[\,]}(\varepsilon, \mathcal{M}^k, \|\cdot\|_p) \asymp \log N(\varepsilon, \mathcal{M}^k, \|\cdot\|_p) \asymp \varepsilon^{-1/k}.$$

As a special case, it was proved in [13] and [17] that, if $\mathcal{C}([0,1])$ is the collection of convex functions on $[0,1]$ that are bounded by 1, then for all $1 \leq p < \infty$, there exist constants $c(p)$ and $C(p)$ depending only on p, such that for all $0 < \varepsilon < 1$,

$$c(p)\varepsilon^{-1/2} \leq \log N(\varepsilon, \mathcal{C}([0,1]), \|\cdot\|_2) \leq C(p)\varepsilon^{-1/2}. \tag{5.1}$$

We will use this result in the later proof.

A multivariate function $f\colon \mathbb{R}^d \to \mathbb{R}$ is called *separately monotone* or block monotone if it is monotone in each variable while holding the other variables fixed. The metric entropy of bounded separately monotone functions on $[0,1]^d$ was studied in [16]. If we denote by $\mathcal{M}_d$ the class of separately monotonic functions on $[0,1]^d$ which are bounded by 1, then roughly speaking it is proved in [16] that for all $1 \leq p < \infty$

$$\log N_{[\,]}(\varepsilon, \mathcal{M}_d, \|\cdot\|_p) \asymp \log N(\varepsilon, \mathcal{M}_d, \|\cdot\|_p) \asymp \varepsilon^{-\max\{d,(d-1)p\}}. \quad (5.2)$$

On the other hand, if we let $\mathcal{C}_d(\Omega)$ be the class of convex functions on $\Omega \subset \mathbb{R}^d$ that are bounded by 1, then it was proved by Guntuboyina and Sen [18] for $\Omega = [0,1]^d$, and by Gao and Wellner [15] for all convex polytopes that for all $1 \leq p < \infty$

$$\log N_{[\,]}(\varepsilon, \mathcal{C}_d(\Omega), \|\cdot\|_p) \asymp \log N(\varepsilon, \mathcal{C}_d(\Omega), \|\cdot\|_p) \asymp \varepsilon^{-d/2}.$$

Furthermore, it was proved by Gao and Wellner [15] that for all $1 \leq p < \infty$

$$\log N(\varepsilon, \mathcal{C}_d(D), \|\cdot\|_p) \geq C\varepsilon^{-\beta},$$

where $\beta = \max\{(d-1)p/2, d/2\}$.

5.2 STATEMENTS OF THE MAIN RESULTS

Unlike separately monotonicity which is an easy condition to check, convexity of a function is typically difficult to check. It is natural to consider functions which are convex or concave in each variable while holding the other variables fixed. We call such functions *separately con functions*. In this chapter, I will prove that

Theorem 43. *Let $\mathcal{F}([0,1]^d)$ denote the collection of separately con functions on $[0,1]^d$ that are bounded by 1. Then, for any $1 \leq p < \infty$, there exists a constant $C(d,p)$ depending only on d and p, such that for all $0 < \varepsilon < 1$,*

$$\log N(\varepsilon, \mathcal{F}([0,1]^d), \|\cdot\|_p) \leq C(d,p)\varepsilon^{-d+\frac{1}{2}}. \quad (5.3)$$

In light of the attention being paid to classes whose functions are defined by generalized shape constraints, separate con is a natural consideration. In this sense, Theorem 43 complements those of [16], [18] and [15] mentioned above.

Remark 44. *I conjecture that in the case $d = 2$, the rate $\varepsilon^{-3/2}$ is optimal.*

5.3 PROOF OF THE MAIN RESULT

We prove the theorem by induction on the dimension d. When $d = 1$, the theorem reduces to equation (5.1). So, the theorem is true for $d = 1$, with $C(1, p) = C(p)$, where $C(p)$ is the constant appear in equation (5.1).

Now, assume that Theorem 43 holds for $d \in \mathbb{N}$. We show that it holds for $d + 1$ with constant

$$C(d+1, p) = \frac{4^d 8^{p+1} C(d, p)}{1 - 2^{-d-\frac{1}{2}}}.$$

We first note that since for any $f \in \mathcal{F}([0,1]^{d+1})$, we have $\|f - 0\|_p \le 1$, we have

$$N(\varepsilon, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p) = 1.$$

for all $\varepsilon \ge 1$. Thus, the theorem is clearly true if $\varepsilon \ge 1$. Now, we assume the induction hypothesis holds for some $n \in \mathbb{N}$, that is, that the theorem is true for all $\varepsilon \ge 2^{-n}$, that is,

$$\log N(\varepsilon, \mathcal{F}([0,1]^{d+1}), \|\cdot\|) \le C(d+1, p)\varepsilon^{-(d+1)+\frac{1}{2}}, \tag{5.4}$$

for all $\varepsilon \ge 2^{-n}$. We show that it is also true for $\varepsilon \ge 2^{-n-1}$.

We first choose $\theta = 8^{-p}$, and define

$$D_\theta := (\theta, 1-\theta) \times [0,1]^d,$$

$$B_1 := [0, \theta] \times [0,1]^d,$$

and

$$B_2 := [\theta, 1-\theta] \times [0,1]^d.$$

Then, for each $f \in \mathcal{F}([0,1]^{d+1})$, we can write

$$f = \mathbb{1}_{D_\theta} f + \mathbb{1}_{B_1} f + \mathbb{1}_{B_2} f. \tag{5.5}$$

If we define

$$\mathcal{G}_1 = \left\{ \mathbb{1}_{B_1} f \;\middle|\; f \in \mathcal{F}([0,1]^{d+1}) \right\},$$
$$\mathcal{G}_2 = \left\{ \mathbb{1}_{B_2} f \;\middle|\; f \in \mathcal{F}([0,1]^{d+1}) \right\},$$

and

$$\mathcal{F}_\theta = \left\{ \mathbb{1}_{D_\theta} f \;\middle|\; f \in \mathcal{F}([0,1]^{d+1}) \right\},$$

then, we have

$$\mathcal{F}([0,1]^{d+1}) \subset \mathcal{G}_1 + \mathcal{G}_2 + \mathcal{F}_\theta.$$

If $\mathcal{N}_1$, and $\mathcal{N}_2$ are $\varepsilon/4$-nets of $\mathcal{G}_1$ and $\mathcal{G}_2$ with minimum cardinality, and if $\mathcal{N}_\theta$ is an $\varepsilon/2$-net of $\mathcal{F}_\theta$ with minimum cardinality, then, $\mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_\theta$ is an ε-net of $\mathcal{F}([0,1]^{d+1}$. Indeed, for any $f \in \mathcal{F}([0,1]^{d+1})$, $\mathbb{1}_{B_1} f \in \mathcal{G}_1$, so there exists $g_1 \in \mathcal{N}_1$ such that

$$\|\mathbb{1}_{B_1} f - g_1\|_p \leq \varepsilon/4.$$

Similarly, there exist $g_2 \in \mathcal{N}_2$ and $g_\theta \in \mathcal{N}_\theta$ such that

$$\|\mathbb{1}_{B_2} f - g_2\|_p \leq \varepsilon/4;$$

$$\|\mathbb{1}_{D_\theta} f - g_\theta\|_p \leq \varepsilon/2.$$

Consequently,

$$\|f - (g_1 + g_2 + g_\theta)\|_p \leq \|\mathbb{1}_{B_1} f - g_1\|_p + \|\mathbb{1}_{B_2} f - g_2\|_p + \|\mathbb{1}_{D_\theta} f - g_\theta\|_p \leq \varepsilon.$$

Therefore,

$$N(\varepsilon, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p) \leq N(\varepsilon/4, \mathcal{G}_1, \|\cdot\|_p) \cdot N(\varepsilon/4, \mathcal{G}_2, \|\cdot\|_p) \cdot N(\varepsilon/2, \mathcal{F}_\theta, \|\cdot\|_p). \tag{5.6}$$

The following lemma enables us to estimate the covering number for $\mathcal{G}_1$ and $\mathcal{G}_2$.

Lemma 45. *For any $\eta > 0$,*

$$N(\eta, \mathcal{G}_1, \|\cdot\|_p) = N(8\eta, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p); \tag{5.7}$$

$$N(\eta, \mathcal{G}_2, \|\cdot\|_p) = N(8\eta, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p). \tag{5.8}$$

Proof. Let $f_1, \dots, f_N$ be a 8η-net of $\mathcal{F}([0,1]^{d+1})$. We define

$$g_i(t, x_2, \dots, x_{d+1}) = f_i(\theta^{-1}t, x_2, \dots, x_{d+1})$$

on $[0,\theta] \times [0,1]^d$, $1 \leq i \leq N$. We claim that $g_1, \dots, g_N$ is an η-net of $\mathcal{G}_1$. Indeed, for any $g \in \mathcal{G}_1$, the function

$$f(x_1, x_2, \dots, x_{d+1}) := g(\theta x_1, x_2, \dots, x_{d+1})$$

belongs to $\mathcal{F}([0,1]^{d+1})$. Thus, there exists some f_i such that $\|f - f_i\|_p \leq 2\eta$. That is

$$\begin{aligned}(8\eta)^p \geq & \int_{[0,1]^d} \int_0^1 |f(x_1, x_2, \dots, x_{d+1}) - f_i(x_1, x_2, \dots, x_{d+1})|^p dx_1 dx_2 \cdots dx_{d+1} \\ = & \int_{[0,1]^d} \int_0^1 |g(\theta x_1, x_2, \dots, x_{d+1}) - g_i(\theta x_1, x_2, \dots, x_{d+1})|^p dx_1 dx_2 \cdots dx_{d+1} \\ = & \theta^{-1} \int_{[0,1]^d} \int_0^\theta |g(t, x_2, \dots, x_{d+1}) - g_i(t, x_2, \dots, x_{d+1})|^p dt\, dx_2 \cdots dx_{d+1} \\ = & 8^p \|g - g_i\|_p^p,\end{aligned}$$

which implies that $\|g - g_i\|_p \leq \eta$. Thus,

$$N(\eta, \mathcal{G}_1, \|\cdot\|_p) \leq N(8\eta, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p).$$

To prove the other direction, we let $g_1, \dots, g_M$ be an η-net of $\mathcal{G}_1$ with minimum cardinality. We define

$$f_i(x_1, x_2, \dots, x_{d+1}) = g_i(\theta^{-1}x_1, x_2, \dots, x_{d+1}).$$

We claim that $f_1, \dots, f_N$ is a 2η-net of $\mathcal{F}([0,1]^{d+1})$. Indeed, for any $f \in \mathcal{F}([0,1]^{d+1})$, since the function

$$g(t, x_2, \dots, x_{d+1}) := f(\theta t, x_2, \dots, x_{d+1})$$

belongs to $\mathcal{G}_1$, there exists g_i, such that $\|g - g_i\|_p \le \eta$. That is,

$$\begin{aligned}
\eta^p \ge & \int_{[0,1]^d} \int_0^\theta |g(t, x_2, \dots, x_{d+1}) - g_i(t, x_2, \dots, x_{d+1})|^p dt dx_2 \cdots dx_{d+1} \\
= & \int_{[0,1]^d} \int_0^\theta |f(\theta^{-1}t, x_2, \dots, x_{d+1}) - f_i(\theta^{-1}t, x_2, \dots, x_{d+1})|^p dt dx_2 \cdots dx_{d+1} \\
= & \theta \int_{[0,1]^d} \int_0^1 |f(x_1, x_2, \dots, x_{d+1}) - f_i(x_1, x_2, \dots, x_{d+1})|^p dx_1\, dx_2 \cdots dx_{d+1} \\
= & 8^{-p} \|f - f_i\|_p^p,
\end{aligned}$$

which implies that $\|f - f_i\|_p \le 8\eta$. Thus,

$$N(\eta, \mathcal{G}_1, \|\cdot\|_p) \ge N(8\eta, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p).$$

This proves (5.7). The proof of (5.8) is similar, and we thus omit it. □

Now, applying Lemma 45 and the induction assumption (5.4), we obtain that for all $\varepsilon \ge 2^{-n-1}$,

$$\log N(\frac{\varepsilon}{4}, \mathcal{G}_1, \|\cdot\|_p) = \log N(2\varepsilon, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p) \le C(d+1, p)(2\varepsilon)^{-(d+1)-\frac{1}{2}}; \quad (5.9)$$

$$\log N(\frac{\varepsilon}{4}, \mathcal{G}_1, \|\cdot\|_p) = \log N(2\varepsilon, \mathcal{F}([0,1]^{d+1}), \|\cdot\|_p) \le C(d+1, p)(2\varepsilon)^{-(d+1)-\frac{1}{2}}. \quad (5.10)$$

Now, we turn to estimating the covering number of $\mathcal{F}_\theta$. To this end, we first prove that for any $f \in \mathcal{F}_\theta$,

$$\left|\frac{\partial f}{\partial x_1}\right| \leq \frac{2}{\theta} = 2 \cdot 8^p. \tag{5.11}$$

Indeed, for each $(x_2, \ldots, x_{d+1}) \in [0,1]^d$, the function $f(t, x_2, \ldots, x_{d+1})$ is convex or concave on $t \in [0,1]$, and bounded by 1. Thus, setting

$$A := \frac{|f(\theta, x_2, \ldots, x_{d+1}) - f(0, x_2, \ldots, x_{d+1})|}{\theta},$$

and

$$B := \frac{|f(1, x_2, \ldots, x_{d+1}) - f(1-\theta, x_2, \ldots, x_{d+1})|}{\theta},$$

we have for any $t, s \in (\theta, 1-\theta)$,

$$\begin{aligned} &\frac{|f(t, x_2, \ldots, x_{d+1}) - f(s, x_2, \ldots, x_{d+1})|}{|t-s|} \\ \leq &\max\{A, B\} \\ \leq &\frac{2}{\theta} \\ = &2 \cdot 8^p, \end{aligned}$$

which implies (5.11).

Now, we let

$$\theta_k = \theta + k \cdot 8^{-p-1}\varepsilon$$

for $k = 1, \ldots, m$, where m is the largest integer such that $\theta_m \leq 1 - \theta$. Clearly, we have $m \leq 8^{p+1}\varepsilon^{-1}$. For any $t \in (\theta, 1-\theta)$, we can find some θ_k such that $|t - \theta_k| \leq 8^{-p-1}\varepsilon$. By (5.11), for any $(x_2, \ldots, x_{d+1}) \in [0,1]^d$, we have

$$|f(t, x_2, \ldots, x_{d+1}) - f(\theta_k, x_2, \ldots, x_{d+1})| \leq 2 \cdot 8^p \cdot 8^{-p-1}\varepsilon = \frac{\varepsilon}{4}.$$

For each $k = 1, \dots, m$, consider the function

$$f_k(x_2, \dots, x_{d+1}) := f(\theta_k, x_2, \dots, x_{d+1}).$$

As a function on $[0,1]^d$, $f_k \in \mathcal{F}([0,1]^d)$. By the induction hypothesis, there exists an $\varepsilon/4$-net $\mathcal{S}_k$ of minimum cardinality

$$N(\varepsilon/4, \mathcal{F}([0,1]^d), \|\cdot\|_p) \le e^{C(d,p)(\varepsilon/4)^{-d+\frac{1}{2}}},$$

such that for some $g_k \in \mathcal{S}_k$, $\|f_k - g_k\|_p \le \varepsilon/4$, that is,

$$\left(\int_{[0,1]^d} |f(\theta_k, x_2, \dots, x_{d+1}) - g(x_2, \dots, x_{d+1})|^p dx_2 \cdots dx_{d+1} \right)^{1/p} \le \frac{\varepsilon}{4}.$$

Now, we define

$$\begin{aligned} g(t, x_2, \dots, x_{d+1}) =& \mathbb{1}_{(\theta,\theta_2)}(t) g_1(x_2, \dots, x_{d+1}) + \sum_{k=2}^{m-1} \mathbb{1}_{[\theta_k,\theta_{k+1})}(t) g_k(x_2, \dots, x_{d+1}) \\ &+ \mathbb{1}_{[\theta_m, 1-\theta)}(t) g_m(x_2, \dots, x_{d+1}), \end{aligned}$$

and

$$\begin{aligned} \widetilde{f}(t, x_2, \dots, x_{d+1}) =& \mathbb{1}_{(\theta,\theta_2)}(t) f(\theta_1, x_2, \dots, x_{d+1}) + \sum_{k=2}^{m-1} \mathbb{1}_{[\theta_k,\theta_{k+1})}(t) f(\theta_k, x_2, \dots, x_{d+1}) \\ &+ \mathbb{1}_{[\theta_m, 1-\theta)}(t) f(\theta_m, x_2, \dots, x_{d+1}). \end{aligned}$$

Then, we have

$$|f(t, x_2, \dots, x_{d+1}) - \widetilde{f}(t, x_2, \dots, x_{d+1})| \le \frac{\varepsilon}{4}.$$

Thus,

$$\|f - g\| \le \frac{\varepsilon}{4} + \|\widetilde{f} - g\|_p. \tag{5.12}$$

Since

$$\begin{aligned}\|\widetilde{f}-g\|_p^p &= \int_{\theta}^{\theta_2}\int_{[0,1]^d}|f(\theta_1,x_2,\dots,x_{d+1})-g_1(x_2,\dots,x_{d+1})|^p dx_2\cdots dx_{d+1}dt\\ &\quad+\sum_{k=2}^{m-1}\int_{\theta_k}^{\theta_{k+1}}\int_{[0,1]^d}|f(\theta_k,x_2,\dots,x_{d+1})-g_k(x_2,\dots,x_{d+1})|^p dx_2\cdots dx_{d+1}dt\\ &\quad+\int_{\theta_m}^{1-\theta}|f(\theta_m,x_2,\dots,x_{d+1})-g_m(x_2,\dots,x_{d+1})|^p dx_2\cdots dx_{d+1}dt\\ &\le \int_{\theta}^{1-\theta}\left(\frac{\varepsilon}{4}\right)^p dt\\ &\le \left(\frac{\varepsilon}{4}\right)^p,\end{aligned}$$

plugging into (5.12), we obtain

$$\|f-g\|_p \le \frac{\varepsilon}{2}.$$

Since there are no more than

$$\left[N(\varepsilon/4,\mathcal{F}([0,1]^d),\|\cdot\|_p)\right]^m \le e^{mC(d,p)(\varepsilon/4)^{-d+\frac{1}{2}}} \le e^{4^d 8^{p+1}C(d,p)\varepsilon^{-d-\frac{1}{2}}},$$

realizations of g, we obtain

$$\log N(\varepsilon/2,\mathcal{F}_\theta,\|\cdot\|_p) \le 4^d 8^{p+1}C(d,p)\varepsilon^{-d-\frac{1}{2}}.$$

Plugging this and (5.9) and (5.10) into (5.6), we obtain that for all $\varepsilon \ge 2^{-n-1}$,

$$\begin{aligned}\log N(\varepsilon,\mathcal{F}([0,1]^{d+1}),\|\cdot\|_p) \le& 2C(d+1,p)(2\varepsilon)^{-(d+1)-\frac{1}{2}} + 4^d 8^{p+1}C(d,p)\varepsilon^{-d-\frac{1}{2}}\\ \le& C(d+1,p)\varepsilon^{-(d+1)+\frac{1}{2}},\end{aligned}$$

where in the last inequality we used the relation that

$$C(d+1,p)=\frac{4^d 8^{p+1}C(d,p)}{1-2^{-d-\frac{1}{2}}}.$$

Hence, by mathematical induction, (5.3) is true for dimension $d+1$ and all $\varepsilon > 0$. This finishes the proof of Theorem 43.

Bibliography

[1] M. Š. Birman and M. Z. Solomjak. Piecewise polynomial approximations of functions of classes ${W_p}^{\alpha}$. *Mat. Sb. (N.S.)*, 73 (115):331–355, 1967.

[2] R. Blei, F. Gao, and W. V. Li. Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.*, 135(12):4009–4018, 2007.

[3] E. M. Bronšteĭn. ε-entropy of convex sets and functions. *Sibirsk. Mat. Ž.*, 17(3):508–514, 715, 1976.

[4] B. Carl. A remark on a paper: "Entropy numbers for noncompact selfadjoint operators in Hilbert spaces" [Math. Nachr. **100** (1981), 213–219; MR 82k:47033] by D. E. Edmunds and H. Triebel. *Math. Nachr.*, 103:39–43, 1981.

[5] B. Carl, A. Hinrichs, and A. Pajor. Gelfand numbers and metric entropy of convex hulls in Hilbert spaces. *Positivity*, 17(1):171–203, 2013.

[6] B. Carl, I. Kyrezi, and A. Pajor. Metric entropy of convex hulls in Banach spaces. *J. London Math. Soc. (2)*, 60(3):871–896, 1999.

[7] A. Colesanti and P. Manselli. Geometric and isoperimetric properties of sets of positive reach in E^d. *Atti Semin. Mat. Fis. Univ. Modena Reggio Emilia*, 57:97–113 (2011), 2010.

[8] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[9] D. E. Edmunds and H. Triebel. *Function spaces, entropy numbers, differential operators*, volume 120 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[10] V. D. Erohin. Best linear approximation of functions analytically continuable from a given continuum to a given region. *Uspehi Mat. Nauk*, 23(1 (139)):91–132, 1968.

[11] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[12] F. Gao. Entropy of absolute convex hulls in Hilbert spaces. *Bull. London Math. Soc.*, 36(4):460–468, 2004.

[13] F. Gao. Entropy estimate for k-monotone functions via small ball probability of integrated Brownian motion. *Electron. Commun. Probab.*, 13:121–130, 2008.

[14] F. Gao. Optimality of CKP-inequality in the critical case. *Proc. Amer. Math. Soc.*, 142(3):909–914, 2014.

[15] F. Gao and J. A. Wellner. Metric Entropy of Convex Functions on $\mathbb{R}^n$ (preprint).

[16] F. Gao and J. A. Wellner. Entropy estimate for high-dimensional monotonic functions. *J. Multivariate Anal.*, 98(9):1751–1764, 2007.

[17] F. Gao and J. A. Wellner. On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Sci. China Ser. A*, 52(7):1525–1538, 2009.

[18] A. Guntuboyina and B. Sen. Covering numbers for convex functions. *IEEE Trans. Inform. Theory*, 59(4):1957–1965, 2013.

[19] M. Guthe, P. Borodin, and R. Klein. Fast and Accurate Hausdorff Distance Calculation Between Meshes. *Journal of WSCG*, 13(2):41–48, February 2005.

[20] Q. Han and J. A. Wellner. Multivariate Convex Regression: Global Risk Bounds and Adapatation (preprint).

[21] A. N. Kolmogoroff. Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Ann. of Math. (2)*, 37(1):107–110, 1936.

[22] A. N. Kolmogorov. On certain asymptotic characteristics of completely bounded metric spaces. *Dokl. Akad. Nauk SSSR (N.S.)*, 108:385–388, 1956.

[23] A. N. Kolmogorov. On linear dimensionality of topological vector spaces. *Dokl. Akad. Nauk SSSR*, 120:239–241, 1958.

[24] A. N. Kolmogorov and V. M. Tihomirov. ε-entropy and ε-capacity of sets in functional space. *Amer. Math. Soc. Transl. (2)*, 17:277–364, 1961.

[25] J. N. McDonald and N. A. Weiss. *A course in real analysis*. Academic Press, Inc., San Diego, CA, 1999. Biographies by Carol A. Weiss.

[26] B. S. Mitjagin. Approximate dimension and bases in nuclear spaces. *Uspehi Mat. Nauk*, 16(4 (100)):63–132, 1961.

[27] A. Pinkus. *n-widths in approximation theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1985.

[28] S. A. Smoljak. ε-entropy of the classes $E_{s^{\alpha},k}(B)$ and $W_{s^{\alpha}}(B)$ in the metric of L_2. *Soviet Math. Dokl.*, 1:192–196, 1960.

[29] V. N. Temljakov. Approximation of functions from the Lipschitz class by algebraic polynomials. *Mat. Zametki*, 29(4):597–602, 634, 1981.

[30] C. Thäle. 50 years sets with positive reach—a survey. *Surv. Math. Appl.*, 3:123–165, 2008.

[31] A. F. Timan. *Theory of approximation of functions of a real variable*. Dover Publications, Inc., New York, 1994. Translated from the Russian by J. Berry, Translation edited and with a preface by J. Cossar, Reprint of the 1963 English translation.

[32] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[33] A. G. Vituškin. *Otsenka slozhnostn zadachi tabulirovaniya*. Sovremennye Problemy Matematiki. Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1959.

[34] H. Widom. Rational approximation and n-dimensional diameter. *J. Approximation Theory*, 5:343–361, 1972. Collection of articles dedicated to J. L. Walsh on his 75th birthday, IV (Proc. Internat. Conf. Approximation Theory, Related Topics and Appl., Univ. Maryland, College Park, Md., 1970).