Metric Entropy of Convex Hulls

Fuchang Gao

University of Idaho

Abstract

Let T be a precompact subset of a Hilbert space. The metric entropy of the convex hull of T is estimated in terms of the metric entropy of T, when the latter is of order $\alpha = 2$. The estimate is best possible. Thus, it answers a question left open in [LL] and [CKP].

0.1 Introduction

Let H be a separable Hilbert space and let T be a precompact subset. Define the covering number

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

where $B(x, \varepsilon)$ is the open ε-ball centered at $x \in H$. The set

$$N_\varepsilon(T) := \{t_1, t_2, \dots, t_n\}$$

is called an ε-net of T. The quantity $\log N(T, \varepsilon)$ plays an important role in the theory of empirical processes (cf.[D]). It is called the metric entropy of T.

Let cov(T) denote the convex hull of T. It is natural to ask for good estimates of $\log N(\mathrm{cov}(T), \varepsilon)$ in terms of $\log N(T, \varepsilon)$. It is known (cf. [C]) that if $\log N(T,\varepsilon) < c \cdot \varepsilon^{-\alpha}$ for some $\alpha > 0$, then

$$\begin{array}{ll} \log N(\mathrm{cov}(T),\varepsilon) \le c \cdot \varepsilon^{-2}(\log \varepsilon^{-1})^{1-2/\alpha}, & 0 < \alpha < 2, \\ \log N(\mathrm{cov}(T),\varepsilon) \le c \cdot \varepsilon^{-\alpha}, & \alpha > 2, \end{array}$$

and those are best possible. As we can see from the above that the situation is completely different for $\alpha < 2$ and $\alpha > 2$. The case $\alpha = 2$ was open. In [LL], Li and Linde studied the metric entropy of $\mathrm{cov}(T)$ via certain quantities originated in the theory of majorizing measures. Among others, they obtained some finer estimates of $\log N(\mathrm{cov}(T),\varepsilon)$, which lead to some important partial results for $\alpha = 2$. For example, the upper bounds for the entropy of $\mathrm{cov}(T)$, $T = \{t_1, t_2, \ldots\}$, $\|t_i\| \le a_i$, by functions of the a_i's only. Their results are optimal for the slowly decreasing sequence (a_i). However, in general, the estimate of the metric entropy of $\mathrm{cov}(T)$ for the case $\alpha = 2$ was left open.

In this paper, we give the best possible estimate for the case $\alpha = 2$. More precisely, we prove the following

Theorem 1 *Let H be a separable Hilbert space and let T be a precompact subset of H.*

(i) *Suppose* $\log N(T,\varepsilon) < \varepsilon^{-2}$, *then for some* $c > 0$,

$$\log N(cov(T),\varepsilon) \le c \cdot \varepsilon^{-2}(\log \varepsilon^{-1})^2;$$

(ii) *There exists a set T, and a constant $c > 0$, such that*

$$\sup_{\varepsilon > 0} \varepsilon^2 \log N(T,\varepsilon) \le 8,$$

and for all $\varepsilon < c$,

$$\log N(cov(T), \varepsilon) \geq c\varepsilon^{-2}(\log(\varepsilon^{-1}))^2.$$

0.2 Proof of (i)

Without loss of generality, we assume the diameter of T is 1. For $k \geq 1$, let N_k be a 2^{-k}-net of T with minimal cardinality. Denote $D_1 = N_1 \cup \{0\}$ and

$$D_n = \{z \in N_n - N_{n-1} : \|z\| \leq 2^{-n+1}\} \cup \{0\}$$

for $n > 1$. Then

$$T \subset D_1 + D_2 + \cdots + D_n + \cdots,$$

where "+" means the Minkowsky sum. By the assumption of (i), D_n consists of no more than $e^{c2^{2n}}$ vectors for some constant $c > 0$. Denote $C_n = \mathrm{cov}(D_n)$ and $E_n = C_1 + C_2 + \cdots + C_n$, then we have

$$\mathrm{cov}(T) \subset C_1 + C_2 + \cdots + C_n + \cdots = E_n + C_{n+1} + \cdots.$$

For any $0 < \varepsilon < 1/4$, suppose $2^{-n+2} \leq \varepsilon < 2^{-n+3}$. Because $C_{n+1} + C_{n+2} + \cdots$ has diameter at most 2^{-n+1}, we have

$$\log N(\mathrm{cov}(T), \varepsilon) \leq \log N(E_n, 2^{-n+1}).$$

To estimate the right side above, we need the following lemma, whose proof is standard.

Lemma 1 *There exists a constant c, such that for any* $\lambda > 0$,

$$\log N(E_n, \lambda) \leq cn^2 \cdot \lambda^{-2}.$$

Proof: For each $k \le n$, suppose $D_k = \{x_1, x_2, \ldots, x_{d_k}\}$, where d_k is the cardinality of D_k. Thus, $d_k \le e^{c2^{2k}}$. For each $z_k \in C_k$, z_k can be expressed as

$$z_k = \sum_{i=1}^{d_k} a_i x_i, \ a_i \ge 0, \ \sum_{i=1}^{d_k} a_i \le 1.$$

Define random vector Z_k, so that

$$\Pr(Z_k = x_i) = a_i, 1 \le i \le d_k, \text{ and } \Pr(Z_k = 0) = 1 - \sum_{i=1}^{d_k} a_i.$$

Let $Z_{k,1}, Z_{k,2}, \ldots, Z_{k,m_k}$ and $Z'_{k,1}, Z'_{k,2}, \ldots, Z'_{k,m_k}$ be independent copies of Z_k. Then

$$E \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} = z_k.$$

Thus, by convexity and symmetrization, we have

$$\begin{aligned}
E \left\| \sum_{k=1}^n z_k - \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} \right\| &= E \left\| E' \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} Z'_{k,i} - \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} \right\| \\
&\le EE' \left\| \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} (Z'_{k,i} - Z_{k,i}) \right\| \\
&= EE' \left\| \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} (Z'_{k,i} - Z_{k,i}) r_{k,i}(t) \right\| \\
&\le 2E \left\| \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} r_{k,i}(t) \right\|
\end{aligned}$$

where $(r_{k,i}(t))$, $1 \le k \le n$, $1 \le i \le m_k$, is a Rademacher sequence. Integrating with respect to t over $[0,1]$, and using Fubini, we obtain

$$\begin{aligned}
E \left\| \sum_{k=1}^n z_k - \sum_{k=1}^n \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} \right\| &\le 2E \left(\sum_{k=1}^n \frac{1}{m_k^2} \sum_{i=1}^{m_k} \|Z_{k,i}\|^2 \right)^{1/2} \\
&\le 2 \left(\sum_{k=1}^n \frac{1}{m_k} 2^{-2k+2} \right)^{1/2} = \lambda,
\end{aligned}$$

taking $m_k = 4n2^{-2k+2}\lambda^{-2}$. This in particular implies that for some realization,

$$\left\| \sum_{k=1}^{n} z_k - \sum_{k=1}^{n} \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} \right\| \leq \lambda.$$

But, there is no more than

$$\prod_{k=1}^{n} (d_k)^{m_k} \leq e^{cn^2\lambda^{-2}}$$

possible realizations of $\sum_{k=1}^{n} \sum_{i=1}^{m_k} Z_{k,i}/m_k$. The lemma follows. □

Applying Lemma 1 with $\lambda = 2^{-n+1}$, and keeping in mind that $2^{-n+2} \leq \varepsilon < 2^{-n+3}$, we obtain

$$\begin{aligned} \log N(\mathrm{cov}(T), \varepsilon) &\leq \log N(E_n, 2^{-n+1}) \\ &\leq c \cdot n^2 2^{-2n+2} \\ &= c' \varepsilon^{-2} (\log \varepsilon^{-1})^2. \end{aligned}$$

Remark 1 *Both Li and Linde pointed out to me that the result (i) can be derived from a result in [CKP]. We include the proof because the current proof seems more transparent, and holds for any Banach space of type 2. Also, it is more convenient to the readers.*

0.3 Proof of (ii)

Let $(\mathbf{e}_k)$ be a standard basis of H. For each integer $k \geq 1$, we define

$$D_k = \left\{ 2^{-k}\mathbf{e}_i : e^{2^{2k-2}} \leq i \leq e^{2^{2k}} \right\} \cup \{0\},$$

and $T = D_1 + D_2 + \cdots + D_k + \cdots$. For any $0 < \varepsilon < 1$, suppose $2^{-n} \leq \varepsilon < 2^{-n+1}$. Define $S_n = D_1 + D_2 + \cdots + D_n$. Because S_n is an 2^{-n}-net of T, and S_n has cardinality no more than

$$\prod_{k \leq n} e^{2^{2k}} \leq e^{2^{2n+1}},$$

we have $\log N(T, \varepsilon) < 2^{2n+1}$. Thus

$$\varepsilon^2 \log N(T, \varepsilon) < 2^{-2n+2} \cdot 2^{2n+1} = 8.$$

To obtain a lower bound for $\log N(\mathrm{cov}(T), \varepsilon)$, we need the following lemma.

Lemma 2 *There exists $c > 0$, such that for $e^{-2^{2k-3}} < \delta < c \cdot 2^{-k}$,*

$$\log N(cov(D_k), \delta) > c \cdot \delta^{-2}.$$

Proof: Denote $I_k = \{i : e^{2^{2k-2}} \leq i < e^{2^{2k}}\}$, and let $|I_k|$ be the cardinality of I_k. Consider the set

$$A = \left\{ \sum_{i \in I_k} a_i \varepsilon \mathbf{e}_i : a_i \text{ is non-negative integer}, \sum_{i \in I_k} a_i \leq 2^{-k}/\varepsilon \right\}.$$

Let m be the largest integer, such that $m \leq 2^{-k}/\varepsilon$. Then A has cardinality no less than $|I_k|^m/m! > |I_k|^{m/2}$. For each $t \in A$, and $2 \leq l < m$, consider

$$B(t, l) = \{s \in A : \|t - s\|_1 \leq l\varepsilon\}.$$

$B(t, l)$ contains no more than $2^l |I_k|^l \leq |I_k|^{2l}$ elements. Thus A contains a subset U of cardinality more than $(|I_k|^{m/2}) \div (|I_k|^{2l})$, whose mutual l_1-

distance between any two elements is at least $l\varepsilon$. Thus, the mutual l_2-distance is at least $\sqrt{l}\varepsilon$. Let $l \approx m/6$. Because $A \subset \text{cov}(D_k)$, we have

$$\begin{aligned}\log N(\text{cov}(D_k), \sqrt{l}\varepsilon) &\leq \log N(\text{co}(D_k), \sqrt{m/6}\cdot\varepsilon)\\ &\geq \log\left(|I_k|^{m/2}/|I_k|^{m/3}\right)\\ &\geq \frac{m}{6}\log|I_k|,\end{aligned}$$

which implies that $\log N(\text{cov}(D_k), \delta) \geq c\cdot\delta^{-2}$ for some $c > 0$ and $e^{-2^{2k-3}} < \delta < c\cdot 2^{-k}$. □

Lemma 3 *For $n \geq 12$, let $m = [n/6]$, and*

$$E_n = cov(D_m) + cov(D_{m+1}) + cov(D_{m+2}) + \cdots + cov(D_n).$$

Then for some constant $c > 0$,

$$\log N(E_n, \sqrt{n}\cdot 2^{-2n-1}) \geq cn\cdot 2^{4n}.$$

Proof: By Lemma 2, for each $m \leq k \leq n$, there exists a set $S_k \subset \text{cov}(D_k)$ of cardinality $L = e^{c\cdot 2^{4n}}$ whose mutual distance between any two elements is at least 2^{-2n}. Consider the set

$$F_n = S_m + S_{m+1} + \cdots + S_n.$$

For $t, s \in F_n$, suppose

$$t = t_m + t_{m+1} + \cdots + t_n, \text{and } s = s_m + s_{m+1} + \cdots + s_n$$

with $t_k \in S_k$ and $s_k \in S_k$. Define the Hamming distance

$$h(t, s) = \text{cardinality of } \{k : t_k \neq s_k, m \leq k \leq n\}.$$

For each $t \in F_n$, the ball

$$B_h(t, n/3) := \{s \in F_n : h(t, s) \leq n/3\}$$

contains no more than $(nL)^{n/4} < L^{n/3}$ elements. Thus F_n contains a subset of cardinality $L^{n-m} \div L^{n/3} \geq L^{n/2}$, whose mutual Hamming distance between any two elements is at least $n/4$. Thus the mutual l_2-distance is at least $\sqrt{n} \cdot 2^{-2n-1}$. This implies that

$$\log N(F_n, \sqrt{n} \cdot 2^{-2n-1}) \geq \frac{n}{2} \log L = \frac{cn}{2} 2^{4n}.$$

□

Now we finish the proof of (ii). For any $0 < \varepsilon < 2^{-24}$, there exists $n \geq 12$, such that

$$\sqrt{n+1} \cdot 2^{-2n-3} < \varepsilon \leq \sqrt{n} \cdot 2^{-2n-1}.$$

Because $F_n \subset \text{cov}(T)$, we have

$$\begin{aligned} \log N(\text{cov}(T), \varepsilon) &\geq \log N(F_n, \sqrt{n} \cdot 2^{-2n-1}) \\ &\geq \frac{cn}{2} 2^{4n} \\ &\geq c' \varepsilon^2 (\log \varepsilon^{-1})^{-2}. \end{aligned}$$

Acknowledgment The author thanks J. Creutzig, W. Li and W. Linde for there interest and valuable comments.

0.4 Reference

[BLL] B. Bühler, W. Li and W. Linde, *Location of Majorizing Measures*, Asymptotic Methods in Probability and Statistics with Applications, Birkhauser, Boston.(to appear)

[BP] K. Ball and A. Pajor, *The entropy of convex bodies with "few" extreme points*, London Math. Soc. Lecture Note Ser. **158**(1990), 25-32.

[C] B. Carl, *Metric entropy of convex hulls in Hilbert spaces*, Bull. London Math. Soc. **29**(1997), 452-458.

[CKP] B. Carl, I. Kyrezi and A. Pajor, *Metric entropy of convex hulls in Banach spaces*, preprint, 1997.

[D] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Funct. Anal. **1**(1967), 290-330.

[LL] W. Li and W. Linde, *Metric Entropy of Convex Hulls in Hilbert Spaces*, Studia Mathematica (to appear).

Fuchang Gao

Department of Mathematics

University of Idaho

Moscow, Idaho 83844-1103

fuchang@uidaho.edu