STUDIA MATHEMATICA 139 (1) (2000)

Metric entropy of convex hulls in Hilbert spaces

by

WENBO V. LI (Newark, DE) and WERNER LINDE (Jena)

Abstract. Let T be a precompact subset of a Hilbert space. We estimate the metric entropy of $\mathrm{co}(T)$, the convex hull of T, by quantities originating in the theory of majorizing measures. In a similar way, estimates of the Gelfand width are provided. As an application we get upper bounds for the entropy of $\mathrm{co}(T)$, $T = \{t_1, t_2, \ldots\}$, $\|t_j\| \le a_j$, by functions of the a_j's only. This partially answers a question raised by K. Ball and A. Pajor (cf. [1]). Our estimates turn out to be optimal in the case of slowly decreasing sequences $(a_j)_{j=1}^{\infty}$.

1. Introduction. Let H be a separable Hilbert space and let $T \subset H$ be a precompact subset. A suitable measure for the size of T are the *covering numbers* defined by

$$N(T,\varepsilon) := \inf\Big\{n \in \mathbb{N} : \exists t_1, \ldots, t_n \in T \text{ such that } T \subseteq \bigcup_{k=1}^{n} B(t_k;\varepsilon)\Big\}$$

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

If $N(T,\varepsilon)$ grows exponentially, it is more convenient to work with the *metric entropy* of T given by

$$H(T,\varepsilon) := \log N(T,\varepsilon).$$

Let $\mathrm{co}(T)$ denote the convex hull of T. Then it is precompact as well, and it is natural to ask for good estimates of $H(\mathrm{co}(T),\varepsilon)$ in terms of $H(T,\varepsilon)$. Such problems play an important role in the theory of empirical processes (cf. [9]). First results were devoted to the case of "small" sets T, i.e. satisfying $N(T,\varepsilon) \le c\varepsilon^{-\alpha}$ for some $\alpha > 0$. In this case we have the optimal estimate

$$H(\mathrm{co}(T),\varepsilon) \le c\varepsilon^{-2\alpha/(2+\alpha)}$$

(cf. [9], [1], [6] and [5], [16] for recent generalizations of this result). Here

---

2000 *Mathematics Subject Classification*: Primary 60G15; Secondary 52A99, 47B06, 46C05.

Key words and phrases: metric entropy, convex hull, majorizing measure, Gaussian process.

Research of W. V. Li supported in part by NSF.

and throughout the paper, c with or without a subscript always denotes a universal positive constant which may be different at each occurrence.

The case of "big" sets $T \subset H$, i.e.

$$H(T,\varepsilon) \le c\varepsilon^{-\alpha} \tag{1.1}$$

for some $\alpha > 0$, requires different techniques and new phenomenons appear. More precisely, as shown in [6], estimate (1.1) implies

$$H(\mathrm{co}(T),\varepsilon) \le \begin{cases} c\varepsilon^{-2}(\log\varepsilon^{-1})^{1-2/\alpha}, & 0<\alpha<2, \\ c\varepsilon^{-\alpha}, & 2<\alpha<\infty, \end{cases}$$

and again, those are best possible. In particular, this tells us that the situation is completely different for $\alpha < 2$ and $\alpha > 2$ (the case $\alpha = 2$ remains open). One possible explanation for this change of quality at $\alpha = 2$ (we do not know of a purely geometric one) is a close relation between metric entropy and Gaussian stochastic processes. More precisely, let $(X_t)_{t\in H}$ be the isonormal Gaussian process on H, i.e.

$$\mathbb{E}X_t = 0 \quad\text{and}\quad \mathbb{E}X_tX_s = \langle t,s\rangle \quad \text{for all } t,s\in H.$$

Recall that one may use the representation

$$X_t = \sum_{k=1}^{\infty} \xi_k\langle t, f_k\rangle \tag{1.2}$$

where $(\xi_k)_{k\ge1}$ is a sequence of i.i.d. standard normal random variables and $(f_k)_{k\ge1}$ is any complete orthonormal system in H. Now, if $T\subset H$ is as before, we may define its *l-width* by

$$l(T) := \sup\{\mathbb{E}\sup_{t\in S}|X_t| : S\subseteq T \text{ finite}\}. \tag{1.3}$$

Then a basic result of R. M. Dudley and V. N. Sudakov (cf. [14]) asserts

$$c_1\sup_{\varepsilon>0}\varepsilon\sqrt{H(T,\varepsilon)} \le l(T) \le c_2\int_0^\infty \sqrt{H(T,\varepsilon)}\,d\varepsilon + \sqrt{2/\pi}\inf_{t\in T}\|t\| \tag{1.4}$$

for some universal $c_1, c_2 > 0$. Since $l(T) = l(\mathrm{co}(T))$, this explains why the metric entropy of $\mathrm{co}(T)$ cannot grow faster than ε^{-2} provided that $\int_0^\infty \sqrt{H(T,\varepsilon)}\,d\varepsilon < \infty$. To obtain sharp bounds for $H(\mathrm{co}(T),\varepsilon)$, the estimates in (1.4) do not suffice. For example, they do not provide any information about $H(\mathrm{co}(T),\varepsilon)$ when the integral in (1.4) is infinite. To overcome these difficulties, the main idea in [6] was to investigate the behavior of the integrals

$$\int_\varepsilon^\infty \sqrt{H(T,\delta)}\,d\delta \quad\text{and}\quad \int_0^\varepsilon \sqrt{H(T,\delta)}\,d\delta \tag{1.5}$$

as $\varepsilon\to0$ (in dependence on whether the integral in (1.4) diverges or converges). Although these techniques led to new and interesting results, they do not imply sharp estimates in the critical case $H(T,\varepsilon)\le c\varepsilon^{-2}$. Recall (cf. [8] or [13]) that (1.4) cannot be improved for general T, i.e. it is impossible to characterize the finiteness of $l(T)$ by the entropy integral in (1.4). Consequently, for general T, the integrals in (1.5) cannot be expected to provide optimal estimates for $H(\mathrm{co}(T),\varepsilon)$. Fortunately, the pioneering work of X. Fernique and M. Talagrand (cf. [17] and [18]) provides us with a purely geometric description of $l(T)$ (majorizing measures). So our objective is to replace the quantities in (1.5) by similar ones derived from the theory of majorizing measures. This leads to finer estimates of $H(\mathrm{co}(T),\varepsilon)$ and allows us to treat some important examples in the critical case $\alpha=2$. Our results are optimal in special situations, yet do not answer the most interesting open question, namely, whether or not $\sup_{\varepsilon>0}\varepsilon^2H(T,\varepsilon)<\infty$ always implies $\sup_{\varepsilon>0}\varepsilon^2H(\mathrm{co}(T),\varepsilon)<\infty$. This is because the left hand side of (1.4) does not characterize the finiteness of $l(T)$ either ([1]).

Besides estimates for the metric entropy we also prove upper bounds for the Gelfand width of $\mathrm{co}(T)$, i.e. we give estimates for the minimal diameter of slices of $\mathrm{co}(T)$ with finite-codimensional subspaces.

In Section 5 we apply our results to sets $T=\{t_1,t_2,\ldots\}$ with $\|t_j\|\le a_j$ for some sequence of a_j's tending to zero monotonically. Here we get direct estimates for the metric entropy of $\mathrm{co}(T)$ in terms of the a_j's. In particular, if $a_j = (\log j)^{-1/2}J(\log j)$ with J slowly varying, this leads to new and best possible estimates of the size of $\mathrm{co}(T)$, shedding some new light on the critical case $\alpha = 2$.

Acknowledgements. The authors would like to acknowledge the support by the RiP program at Oberwolfach where this work originated. Furthermore we are grateful to M. A. Lifshits for helpful remarks concerning the organization of this paper.

2. Notations and their relations. Our first objective is to state the convex hull problem in an analytic form. Given $T\subset H$ bounded, the Banach space $l_1(T)$ consists of summable functions on T, i.e. $a=(\alpha_t)_{t\in T}$ belongs to $l_1(T)$ provided that $\|a\|_1 := \sum_{t\in T}|\alpha_t| < \infty$. Note that then at most countably many of the α_t's are different from zero. Setting

$$u_T(a) := \sum_{t\in T}\alpha_t t, \quad a=(\alpha_t)_{t\in T},$$

defines a bounded operator u_T from $l_1(T)$ into H. Moreover, u_T is compact iff T is a precompact subset of H. For our purposes it is more convenient to

([1]) *Added in proof* (March 2000). The above problem has been solved recently by Fuchang Gao (Univ. of Idaho) as follows: If $\sup_{\varepsilon>0}\varepsilon^2H(T,\varepsilon)<\infty$, then this implies $H(\mathrm{co}(T),\varepsilon)\le c\varepsilon^{-2}(\log\varepsilon^{-1})^2$ and, moreover, for general sets $T\subset H$ this is best possible.

work with $\mathrm{aco}(T)$ instead of $\mathrm{co}(T)$. Here

$$\mathrm{aco}(T)=\Big\{\sum_{k=1}^{n}\alpha_k t_k:\sum_{k=1}^{n}|\alpha_k|\le 1,\ t_k\in T,\ n\in\mathbb{N}\Big\}$$

denotes the absolutely convex (symmetric convex) hull of T in H. Indeed, because of

(2.1) $$\mathrm{aco}(T)\subseteq\{u_T(a):\|a\|_1\le 1\}\subseteq\overline{\mathrm{aco}(T)},$$

the metric entropy of $\mathrm{aco}(T)$ (which is of course greater than that of $\mathrm{co}(T)$) is closely related to the sequence of entropy numbers of u_T and this allows us to use basic properties of those numbers. More precisely, for each $k\in\mathbb{N}$ we define the kth *entropy number* of $T\subset H$ by

(2.2) $$\varepsilon_k(T):=\inf\{\varepsilon>0:N(T,\varepsilon)\le k\}$$

and the kth (*dyadic*) *entropy number* by

$$e_k(T):=\varepsilon_{2^{k-1}}(T).$$

In view of (2.1) it follows that

$$e_k(\mathrm{aco}(T))=e_k(u_T)\quad\text{and}\quad N(\mathrm{aco}(T),\varepsilon)=N(u_T(B_{l_1(T)}),\varepsilon)$$

where $e_k(u_T):=e_k(u_T(B_{l_1(T)}))$ and $B_{l_1(T)}$ denotes the closed unit ball of $l_1(T)$ (cf. [7] or [14] for further properties of entropy numbers of operators).

We shall need still another measure for the size of T. If u is an arbitrary operator between Banach spaces E and F, its kth *Gelfand number* $c_k(u)$ is defined by

$$c_k(u):=\inf\{\|u|_M\|:M\subseteq E\text{ and }\operatorname{codim}(M)<k\}.$$

For $T\subset H$ bounded, let u_T be as above. Then we define the kth *Gelfand width* of T (more precisely of $\mathrm{aco}(T)$) by

$$c_k(T):=c_k(u_T).$$

Observe that this width has a geometric meaning. Namely, if T is finite and n is the dimension of H_n, the space spanned by T, then it is not difficult to see that

$$c_k(T)=\inf\{\mathrm{diam}(\mathrm{aco}(T)\cap F):F\subseteq H_n,\ \dim(F)>n-k\}.$$

So $c_k(T)$ measures the minimal diameter of m-dimensional, $m>n-k$, slices of $\mathrm{aco}(T)$ and for arbitrary precompact T one may use

$$c_k(T)=\sup\{c_k(T_0):T_0\subseteq T\text{ finite}\}.$$

The main properties of Gelfand numbers can be found in [14], Chapter 5.

Finally, we relate $l(T)$ defined in (1.3) to the l-norm of u_T^*. Given an operator v from a Hilbert space H into a Banach space E, its l-*norm* is

defined by

(2.3) $$l(v):=\sup_{n\ge1}\mathbb{E}\Big\|\sum_{j=1}^{n}\xi_j v(f_j)\Big\|_E$$

where $f_1,f_2,\ldots$ is an orthonormal basis in H. Normally, this l-norm is defined by second moments (cf. [14], p. 35), which by Fernique's theorem (cf. [13], Cor. 3.2, p. 59) is equivalent to (2.3). For $T\subset H$ let $u_T:l_1(T)\to H$ be as above. Then the dual operator u_T^* maps H into $l_\infty(T)$ (set of bounded functions on T endowed with the sup-norm) and

$$u_T^*(h)=(\langle t,h\rangle)_{t\in T}$$

for $h\in H$. It is easy to see that $l(u_T^*)=l(T)$.

We now state the basic relations between the quantities defined above. The first result relates entropy numbers to Gelfand widths by the so-called Carl inequality (cf. [3], Thm. 1). For later purposes we formulate a recently proved more general statement (cf. Thm. 1.3 of [6]).

PROPOSITION 2.1. *Let b_k be an increasing sequence of positive numbers such that*

$$b_{2k}\le\gamma b_k,\quad k\in\mathbb{N},$$

for some $\gamma\ge1$. Then there is a constant $\kappa\ge1$ only depending on γ such that for all $T\subset H$ and all $n\in\mathbb{N}$,

$$\sup_{1\le k\le n} b_k e_k(\mathrm{aco}(T))\le\kappa\sup_{1\le k\le n} b_k c_k(T).$$

The next result relates the Gelfand widths of a set T to its l-width. This is a reformulation of a basic result due to A. Pajor and N. Tomczak-Jaegermann (cf. [14], Thm. 5.8) in the language of sets.

PROPOSITION 2.2. *For $T\subset H$ we have*

$$\sup_{k\in\mathbb{N}}\sqrt{k}\,c_k(T)\le c\,l(T).$$

3. Majorizing measures. For later purposes we need an inner, purely geometric description of $l(T)$ as well as of related quantities. In the case of l-width this was done by X. Fernique and M. Talagrand (cf. [17]), yet it does not suffice for our purposes. More precisely, we need quantities which either measure the quality of sets T with $l(T)<\infty$ or quantify the "degree of infinity" for $l(T)=\infty$. To do so, we have to modify the basic ideas in [18] slightly.

Here and later on q always denotes a fixed integer sufficiently large ($q\ge16$ suffices). Given $T\subset H$ precompact, a number $i\in\mathbb{Z}$ is chosen as the largest integer for which $N(T,q^{-i})=1$; it will be fixed as long as T is fixed. Let $J\subseteq\{i,i+1,\ldots\}\subset\mathbb{Z}$ be a finite or infinite interval. Then $\mathcal{A}=$

$\{\mathcal{A}_j\}_{j\in J}$ always denotes a sequence of finite partitions of T with the following properties.

(i) $\mathcal{A}_i = \{T\}$ whenever $i \in J$,
(ii) $\mathcal{A}_{j+1}$ always refines $\mathcal{A}_j$ and
(iii) for each $A \in \mathcal{A}_j$ we have $\mathrm{diam}(A) \le 2q^{-j}$.

A sequence $\mathbf{w} = (w_j)_{j\in J}$ of weights is said to be *adapted* (to $\mathcal{A}$) provided that $w_j : \mathcal{A}_j \to [0,1]$, $w_i \equiv 1$ whenever $i \in J$, and moreover, for each $j \in J$,

$$\sum_{A\in\mathcal{A}_j} w_j(A) \le 1.$$

In the classical case $J = \{i, i+1, \ldots\}$ we define a (possibly infinite) number

$$\Theta_{\mathcal{A},\mathbf{w}}(T) := \sup_{t\in T} \sum_{j=i+1}^{\infty} q^{-j} \sqrt{\log \frac{1}{w_j(A_j(t))}}$$

where $A_j(t)$ is the unique set in $\mathcal{A}_j$ with $t \in A_j(t)$.

Set

$$\Theta(T) := \inf\{\Theta_{\mathcal{A},\mathbf{w}}(T) : \mathcal{A} = \{\mathcal{A}_j\}_{j\ge i},\ \mathbf{w} = (w_j)_{j\ge i}\}.$$

The remarkable result about Gaussian processes can now be formulated as follows (cf. [18]).

THEOREM 3.1. *For any $T \subset H$,*

$$c_1\Theta(T) \le \sup_{\substack{S\subseteq T\\ S \text{ finite}}} \mathbb{E} \sup_{t\in S} X_t \le c_2\Theta(T). \tag{3.1}$$

In particular, we have $l(T) < \infty$ iff $\Theta(T) < \infty$.

Let us give a first generalization of the above construction. Choose now $J = \{i, i+1, \ldots, N\}$ for some $N > i$, i.e. we deal with sequences $\mathcal{A} = \{\mathcal{A}_j\}_{j=i}^N$ of partitions of T and adapted weights $\mathbf{w} = (w_j)_{j=i}^N$. Then we put

$$\Theta^N_{\mathcal{A},\mathbf{w}}(T) := \sup_{t\in T} \sum_{j=i+1}^{N} q^{-j} \sqrt{\log \frac{1}{w_j(A_j(t))}}$$

and

$$\Theta^N(T) := \inf\{\Theta^N_{\mathcal{A},\mathbf{w}}(T) : \mathcal{A} = \{\mathcal{A}_j\}_{j=i}^N,\ \mathbf{w} = (w_j)_{j=i}^N\},$$

where the N indicates that only sequences up to order N are used. The main advantage of this quantity is that it is finite for any precompact T, not only for sets with $l(T) < \infty$.

The following generalization of (1.4) may be proved by similar methods to those used in [12] for infinite sequences of partitions.

PROPOSITION 3.1. *For any $T \subset H$ and any $N > i$ we have*

$$c_1 \sup_{i\le j\le N-1} q^{-j}\sqrt{H(T, q^{-j})} \le \Theta^N(T) \le c_2 \int_{q^{-N-1}}^{\infty} \sqrt{H(T,\varepsilon)}\, d\varepsilon.$$

Next we need a quantity which measures the quality of a set T with $l(T) < \infty$. Let $T \subset H$ and $i \in \mathbb{Z}$ be as before. Given $M > i$, this time J is $\{M+1, \ldots\}$, i.e. $\mathcal{A}$ and $\mathbf{w}$ are of the form $\mathcal{A} = \{\mathcal{A}_j\}_{j>M}$ and $\mathbf{w} = (w_j)_{j>M}$. Then we define

$$\Delta^M_{\mathcal{A},\mathbf{w}}(T) := \sup_{t\in T} \sum_{j=M+1}^{\infty} q^{-j} \sqrt{\log \frac{1}{w_j(A_j(t))}}$$

and

$$\Delta^M(T) := \inf\{\Delta^M_{\mathcal{A},\mathbf{w}}(T) : \mathcal{A} = \{\mathcal{A}_j\}_{j>M},\ \mathbf{w} = (w_j)_{j>M}\}.$$

Observe that $\Delta^M(T) < \infty$ for one (each) $M > i$ iff $l(T) < \infty$.

The next result may be regarded as a counterpart to (1.4) for $\Delta^M(T)$ and can be proved by exactly the same methods as in [12].

PROPOSITION 3.2. *For any $M > i$ and $T \subset H$ we have*

$$c_1 \sup_{j>M} q^{-j}\sqrt{H(T, q^{-j})} \le \Delta^M(T) \le c_2 \int_0^{q^{-M}} \sqrt{H(T,\varepsilon)}\, d\varepsilon.$$

Using standard methods (cf. [12] or [18]) one can construct a probability measure μ on T for which

$$\sup_{t\in T} \int_0^{q^{-M}} \sqrt{\log \frac{1}{\mu(B(t;\delta))}}\, d\delta \le c\Delta^M(T) \tag{3.2}$$

with some universal $c > 0$. Indeed, if $\mathcal{A} = \{\mathcal{A}_j\}_{j>M}$ and $\mathbf{w} = (w_j)_{j>M}$ are admissible partitions and weights with

$$\Delta^M_{\mathcal{A},\mathbf{w}}(T) \le 2\Delta^M(T),$$

then for each $j > M$ and each $A \in \mathcal{A}_j$ we choose points $t_j^A \in A$ and define a measure $\widetilde{\mu}$ on T by

$$\widetilde{\mu} := \sum_{j=M+1}^{\infty} \frac{1}{2^{j-M}} \sum_{A\in\mathcal{A}_j} w_j(A)\delta_{t_j^A}.$$

Normalizing $\widetilde{\mu}$ we get a probability measure μ satisfying (3.2).

Combining (3.2) with Proposition 5.2.6 in [10] we obtain the following useful result.

PROPOSITION 3.3. *For any $M > i$ we have*

$$\sup_{\substack{S\subseteq T\\ S \text{ finite}}} \mathbb{E} \sup_{\substack{t,s\in S\\ \|t-s\|\le q^{-M}}} |X_t - X_s| \le c\Delta^M(T). \tag{3.3}$$

REMARK. Other interesting properties of $\Theta^N(T)$ and $\Delta^M(T)$ will be the subject of a separate paper (cf. [2]). For example, as in the classical case, they are equivalent to some expressions defined by measures on T. Moreover, $l(T) < \infty$ iff $\sup_{N>i} \Theta^N(T) < \infty$, and there exist probabilistic descriptions (similar to that of (3.1)) of the purely geometric quantities $\Theta^N(T)$ and $\Delta^M(T)$.

4. Metric entropy of convex hulls. Let us first treat the case of sets $T \subset H$ with $l(T) = \infty$. Recall that $q > 1$ is the sufficiently large fixed natural number used in the definition of $\Theta^N(T)$ and $i \in \mathbb{Z}$ is the largest integer with $N(T, q^{-i}) = 1$.

THEOREM 4.1. *Let T be a precompact subset of H and suppose, for simplicity, $0 \in T$. Then for $k \in \mathbb{N}$ we have*

$$\sqrt{k}\max\{c_k(T), e_k(\operatorname{aco}(T))\} \le c \inf_{N>i}\{\Theta^N(T) + q^{-N}\sqrt{k}\}. \tag{4.1}$$

Proof. Let us first prove the estimate for the Gelfand width. Fix $N > i$ and let $\mathcal{A} = \{\mathcal{A}_j\}_{j=i}^N$ and $\mathbf{w} = (w_j)_{j=1}^N$ be sequences of partitions and adapted weights. In each set $A \in \mathcal{A}_j$ we choose an element s_A and set

$$T_j := \{s_A : A \in \mathcal{A}_j\}.$$

Next define $s_j : T \to T_j$ by $s_j(t) := s_{A_j(t)}$. Recall that $\mathcal{A}_i = \{T\}$, thus by assumption we may choose $T_i = \{0\}$ and $s_i(t) \equiv 0$. If $(X_t)_{t\in T}$ is the isonormal Gaussian process defined in (1.2), then for $t \in T_N$ one has

$$X_t = \sum_{j=i+1}^{N} [X_{s_j(t)} - X_{s_{j-1}(t)}].$$

A standard chaining argument (cf. [12], proof of Thm. 6.1) now implies

$$l(T_N) = \mathbb{E}\sup_{t\in T_N} |X_t| \le c\Theta^N_{\mathcal{A},\mathbf{w}}(T_N) \le c\Theta^N_{\mathcal{A},\mathbf{w}}(T), \tag{4.2}$$

thus by taking the infimum over all $\mathcal{A}, \mathbf{w}$, from (4.2) we derive

$$l(T_N) \le c\Theta^N(T). \tag{4.3}$$

Now we are in a position to apply Proposition 2.2 and obtain

$$\sqrt{k}\, c_k(T_N) \le c\Theta^N(T) \tag{4.4}$$

for any $k \ge 1$. Next observe that $\operatorname{diam}(A) \le 2q^{-N}$ whenever $A \in \mathcal{A}_N$, hence by the choice of T_N it constitutes a $2q^{-N}$-net of T, i.e.

$$T \subseteq T_N + B(2q^{-N}) \tag{4.5}$$

(here and later on $B(\varepsilon)$ denotes the open ε-ball centered at zero). Then (4.4) implies

$$\begin{aligned}\sqrt{k}\, c_k(T) &\le \sqrt{k}\, c_k(T_N) + 2q^{-N}\sqrt{k} \le c\Theta^N(T) + 2q^{-N}\sqrt{k}\\ &\le c(\Theta^N(T) + q^{-N}\sqrt{k}),\end{aligned}$$

completing the proof in this case.

Next we prove the corresponding estimate for the entropy numbers of $\operatorname{aco}(T)$. First observe that this does not follow from the estimate for $c_k(T)$ via Proposition 2.1 because the right hand side of (4.1) depends on k. With the same notation as in the first part of the proof, (4.5) implies

$$\operatorname{aco}(T) \subseteq \operatorname{aco}(T_N) + B(2q^{-N}),$$

which easily gives

$$e_k(\operatorname{aco}(T)) \le e_k(\operatorname{aco}(T_N)) + 2q^{-N}. \tag{4.6}$$

Using (4.3), by (4.6) and Sudakov's minorization theorem (Proposition 2.2 combined with Proposition 2.1 for $b_k = \sqrt{k}$) we finally obtain

$$\sqrt{k}\, e_k(\operatorname{aco}(T)) \le c\, l(T_N) + 2\sqrt{k}\, q^{-N} \le c(\Theta^N(T) + q^{-N}\sqrt{k}),$$

which completes the proof.

REMARK. Since

$$\Theta^N(T) \le c \int_{q^{-N-1}}^{\infty} \sqrt{H(T,\varepsilon)}\, d\varepsilon, \tag{4.7}$$

Theorem 4.1 implies Proposition 5.2 of [6] in the case of Gelfand widths. On the other hand, in view of

$$\Theta^N(T) \ge c \sup_{i\le j\le N-1} q^{-j}\sqrt{H(T,q^{-j})}$$

the improvement in Theorem 4.1 is subtle and important in some circumstances given in Section 5.

Next we treat the case of sets $T \subset H$ with $\Delta^M(T) \to 0$ as $M \to \infty$. We shall see how the behavior of $\Delta^M(T) \to 0$ provides information about the size of $\operatorname{aco}(T)$.

THEOREM 4.2. *Let T be a precompact subset of a Hilbert space H. Given an integer $M > i$, define $N := N(T, q^{-M})$. Then for all integers $m \ge 1$ we have*

$$\sqrt{m}\, c_{m+N}(T) \le c\Delta^M(T).$$

Consequently, if $k, m \geq 1$, *then*

$$\sqrt{m}\, c_{m+k}(T) \leq c\Delta^{M(k)}(T) \tag{4.8}$$

where $M(k)$ *is the maximal* $M > i$ *for which* $q^{-M} \geq \varepsilon_k(T)$ *with* $\varepsilon_k(T)$ *defined in* (2.2).

Proof. Let $T_M \subseteq T$ be an optimal q^{-M}-net, i.e. $\operatorname{card}(T_M) = N(T, q^{-M}) = N$, and define a mapping $s_M : T \to T_M$ such that $\|t - s_M(t)\| \leq q^{-M}$ for all $t \in T$. If $S_M \subset H$ is given by

$$S_M := \{t - s_M(t) : t \in T\},$$

then $T \subseteq S_M + T_M$, hence by well known properties of Gelfand numbers (cf. [14], p. 61)

$$c_{m+k-1}(T) \leq c_m(S_M) + c_k(T_M)$$

for all $m, k \geq 1$. For $k = N + 1 = \operatorname{card}(T_M) + 1$ we have $c_k(T_M) = 0$, thus in view of Proposition 2.2,

$$\sqrt{m}\, c_{m+N}(T) \leq \sqrt{m}\, c_m(S_M) \leq c\, l(S_M) = c \sup_{\substack{S \subseteq T \\ S \text{ finite}}} \mathbb{E} \sup_{t \in S} |X_t - X_{s_M(t)}|. \tag{4.9}$$

To estimate this further we choose an arbitrary finite subset $S \subseteq T$ and without losing generality we assume $T_M \subseteq S$. Since

$$\{(t, s_M(t)) : t \in S\} \subseteq \{(t, s) : t, s \in S,\ \|t - s\| \leq q^{-M}\},$$

by Proposition 3.3 we obtain

$$\mathbb{E} \sup_{t \in S} |X_t - X_{s_M(t)}| \leq \mathbb{E} \sup_{\substack{t, s \in S \\ \|t-s\| \leq q^{-M}}} |X_t - X_s| \leq c\Delta^M(T).$$

This combined with (4.9) proves

$$\sqrt{m}\, c_{m+N}(T) \leq c\Delta^M(T)$$

as asserted. Finally, (4.8) follows from

$$M(k) = \sup\{M > i : N(T, q^{-M}) \leq k\}$$

and completes the proof.

An application of Proposition 3.1 then implies the following (cf. Proposition 5.3 in [6]).

COROLLARY 4.1. *For all* $k, m \geq 1$ *we have*

$$\sqrt{m}\, c_{m+k}(T) \leq c \int_0^{\varepsilon_k(T)} \sqrt{H(T, \varepsilon)}\, d\varepsilon.$$

A basic ingredient in the proof of Theorem 4.2 was that $c_k(T_M) = 0$ for $k > \operatorname{card}(T_M)$. This is no longer valid for the entropy numbers, so we cannot prove a similar estimate for these numbers by the same methods. Fortunately, in most cases Proposition 2.1 applies and leads to the following.

THEOREM 4.3. *Let* β_k *be a decreasing sequence of positive numbers such that*

$$\beta_k \leq \gamma\beta_{2k} \quad \textit{for some } \gamma \geq 1. \tag{4.10}$$

If

$$\Delta^{M(k)}(T) \leq \beta_k \quad \textit{for all } k \in \mathbb{N}, \tag{4.11}$$

then

$$\sqrt{k}\, e_k(\operatorname{aco}(T)) \leq c\beta_k, \quad k \in \mathbb{N}.$$

Proof. If we combine assumption (4.11) with (4.8), then this implies

$$\sqrt{k}\, c_{2k}(T) \leq c\Delta^{M(k)}(T) \leq c\beta_k.$$

Enlarging the constant $c > 0$, by (4.10) this even yields

$$\sqrt{k}\, c_k(T) \leq c\beta_k$$

for each $k \in \mathbb{N}$. An application of Proposition 2.1 with $b_k = \beta_k^{-1}\sqrt{k}$ completes the proof.

5. Convex hulls of sets with few vectors. Next we want to apply the preceding results to sets $T \subset H$ with $T = \{t_1, t_2, \ldots\}$, satisfying $\|t_j\| \leq a_j$, $j = 1, 2, \ldots$, for some sequence $(a_j)_{j=1}^{\infty}$ tending to zero monotonically. One asks for good upper estimates of $e_k(\operatorname{co}(T))$ in terms of the a_j's. If $a_j = j^{-\alpha}$ for some $\alpha > 0$ (fast decay case), the answer is known in a weak form, namely, $e_k(\operatorname{co}(T)) \leq ck^{-\alpha-1/2}$ (cf. [1] or [9]). Here "weak" means that we do not know of a general estimate of $e_k(\operatorname{co}(T))$ in terms of the a_j's only, valid for any polynomial sequence $(a_j)_{j=1}^{\infty}$ (cf. [5] and [16] for recent progress). We shall state and prove such an explicit formula which is sharp for the slow decay case, i.e. if the a_j's tend to zero in logarithmic order. In particular, we obtain optimal estimates in the critical case $a_j = (\log j)^{-1/2} J(\log j)$ with J slowly varying. For non-constant functions J this could not be handled by previously known methods. Recall that for $\|t_j\| \leq c(\log j)^{-1/2}$ one necessarily has $l(T) < \infty$, thus by Sudakov's minorization

$$e_k(\operatorname{aco}(T)) \leq ck^{-1/2}$$

and this is known to be best possible. Our results rest upon sharp estimates for $\Theta^N(T)$ and $\Delta^M(T)$. Before stating and proving them, let us first mention that we may identify $T = \{t_1, t_2, \ldots\}$ with $\mathbb{N}$ endowed with the metric

$$d(n, m)^2 = \|t_n - t_m\|^2 \leq 2(\|t_n\|^2 + \|t_m\|^2) \leq 2(a_n^2 + a_m^2).$$

For simplicity, let us always assume $1/2 < a_1 \leq 1$, so that the enumeration of partitions starts at $i = 1$.

For $q > 1$ fixed as before, we write $\sigma(k) := \operatorname{card}\{m \in \mathbb{N} : a_m \geq q^{-k}\}$.

PROPOSITION 5.1. *If T is as above, then*

$$\Theta^N(T) \leq c(1 + \sup_{1\leq k\leq N} q^{-k}\sqrt{\log\sigma(k)}). \tag{5.1}$$

Proof. Define a sequence $\mathcal{A} = \{\mathcal{A}_j\}_{j=1}^N$ of partitions of $\mathbb{N}$ by

$$\mathcal{A}_j = \{\{1\}, \ldots, \{\sigma(j)\}, B_j\}$$

where $B_j = \{\sigma(j) + 1, \ldots\}$. Since $\operatorname{diam}(B_j) \leq 2q^{-j}$, $\mathcal{A}$ satisfies the assumptions on partitions made in Section 3. Now for $1 \leq j \leq N$ we define weights w_j by $w_j(B_j) = 1/2$ and

$$w_j(\{m\}) := \frac{q-1}{2q} a_m e^{-K_N^2/a_m^2}, \quad 1 \leq m \leq \sigma(j),$$

where

$$K_N = q \sup_{1\leq k\leq N} q^{-k}\sqrt{\log\sigma(k)}.$$

We have $\sum_{A\in\mathcal{A}_j} w_j(A) \leq 1$ because

$$\sum_{j=1}^{\sigma(N)} a_j e^{-K_N^2/a_j^2} = \sum_{k=1}^N \sum_{q^{-k}\leq a_j<q^{-k+1}} a_j e^{-K_N^2/a_j^2} \leq \sum_{k=1}^N q^{-k+1}\sigma(k) e^{-K_N^2 q^{2k-2}}$$
$$= \sum_{k=1}^N q^{-k+1} \exp(\log\sigma(k) - q^{2k} \sup_{1\leq l\leq N} q^{-2l}\log\sigma(l))$$
$$\leq \sum_{k=1}^N q^{-k+1} \leq q/(q-1).$$

So by definition

$$\Theta^N(T) \leq \sup_{m\in\mathbb{N}} \sum_{j=1}^N q^{-j}\sqrt{\log\frac{1}{w_j(A_j(m))}}.$$

For $m > \sigma(N)$ we have

$$\sum_{j=1}^N q^{-j}\sqrt{\log\frac{1}{w_j(A_j(m))}} \leq \frac{q}{q-1}\sqrt{\log 2},$$

while for $m \leq \sigma(N)$,

$$\sum_{j=1}^N q^{-j}\sqrt{\log\frac{1}{w_j(A_j(m))}} \leq \sum_{q^{-j}\leq a_m} q^{-j}\sqrt{\log\frac{1}{w_j(\{m\})}} + \frac{q}{q-1}\sqrt{\log 2}.$$

Note that for $m \leq \sigma(j)$ or, equivalently, $q^{-j} \leq a_m$, we have

$$\sqrt{\log\frac{1}{w_j(\{m\})}} \leq \sqrt{\log\frac{2q}{q-1}} + \sqrt{\log\frac{1}{a_m}} + \frac{K_N}{a_m}$$

and

$$\sum_{q^{-j}\leq a_m} q^{-j} \leq c a_m,$$

so that we finally obtain (recall that we assume $a_1 \leq 1$)

$$\sup_{m\in\mathbb{N}} \sum_{j=1}^N q^{-j}\sqrt{\log\frac{1}{w_j(A_j(m))}} \leq c(1 + K_N),$$

which proves (5.1) by definition of K_N.

REMARK. Observe that by Proposition 3.1,

$$\sup_{1\leq k\leq N-1} q^{-k}\sqrt{H(T, q^{-k})} \leq c\Theta^N(T)$$

for any precompact subset $T \subset H$. Thus (5.1) is asymptotically optimal if T is an orthogonal set with $\|t_j\| = a_j$, which implies $H(T, q^{-k}) \geq \log\sigma(k)$. Moreover, one should compare (5.1) with the weaker estimate

$$\Theta^N(T) \leq c\sum_{j=1}^N q^{-j}\sqrt{\log\sigma(j)},$$

which follows from (4.7).

Next we investigate $\Delta^M(T)$ for T as above.

PROPOSITION 5.2. *Let T be a subset of H as before. Then*

$$\Delta^M(T) \leq c \sup_{l\geq M} q^{-l}\sqrt{\log\sigma(l)}. \tag{5.2}$$

Proof. Define $\mathcal{A}_{M+1}, \mathcal{A}_{M+2}, \ldots$ as before, that is, $\mathcal{A}_j := \{\{1\}, \ldots, \{\sigma(j)\}, B_j\}$. Now for $j > M$ we construct weights w_j by

$$w_j(\{m\}) := \begin{cases} \dfrac{1}{3\sigma(M)} & \text{for } 1 \leq m \leq \sigma(M), \\ \dfrac{q-1}{3q^2} \cdot q^M a_m e^{-C_M^2/a_m^2} & \text{for } \sigma(M) < m \leq \sigma(j), \end{cases}$$
$$w_j(B_j) := 1/3,$$

where

$$C_M := q \sup_{l\geq M} q^{-l}\sqrt{\log\sigma(l)}.$$

We have $\sum_{A\in\mathcal{A}_j} w_j(A) \le 1$ because

$$\sum_{j\ge\sigma(M)} a_j e^{-C_M^2/a_j^2} = \sum_{k=M}^{\infty} \sum_{q^{-k}\le a_j<q^{-k+1}} a_j e^{-C_M^2/a_j^2} \le \sum_{k=M}^{\infty} q^{-k+1}\sigma(k)e^{-C_M^2 q^{2k-2}} \le \sum_{k=M}^{\infty} q^{-k+1} = \frac{q^2}{q-1}\cdot q^{-M}.$$

By the construction

$$\sum_{j=M+1}^{\infty} q^{-j}\sqrt{\log\frac{1}{w_j(A_j(m))}} \le cq^{-M}\sqrt{\log\sigma(M)}$$

for $1\le m\le\sigma(M)$. If $m>\sigma(M)$, then

$$\sum_{j=M+1}^{\infty} q^{-j}\sqrt{\log\frac{1}{w_j(A_j(m))}} \le \sum_{q^{-j}\le a_m} q^{-j}\sqrt{\log\frac{1}{w_j(\{m\})}} + \sum_{j=M+1}^{\infty} q^{-j}\sqrt{\log 3}$$
$$\le \sum_{q^{-j}\le a_m} q^{-j}\left(\sqrt{\log\frac{3q^2}{q-1}} + \sqrt{\log\frac{1}{a_m q^M}} + \frac{C_M}{a_m}\right) + cq^{-M}$$
$$\le c\left(q^{-M} + C_M + a_m\sqrt{\log\frac{1}{a_m q^M}}\right) \le c(C_M + q^{-M})$$

where we used $a_m \le q^{-M}$. This completes the proof by definition of C_M.

REMARK. As above, by Proposition 3.2 estimate (5.2) is optimal for orthogonal $T=\{t_1,t_2,\ldots\}$ with $\|t_j\|=a_j$.

Now we are in a position to prove the announced concrete estimates for the entropy of sets generated by a countable number of vectors in a Hilbert space.

THEOREM 5.1. *Let $T=\{t_1,t_2,\ldots\}\subset H$ and assume that $\|t_j\|\le a_j$, $j=1,2,\ldots$, for some sequence $(a_j)_{j=1}^{\infty}$ of real numbers tending monotonically to zero.*

1. *If $(\log j)^{1/2}a_j$ is increasing, then*

$$\max\{c_k(T), e_k(\operatorname{aco}(T))\} \le ca_{2^k}. \tag{5.3}$$

2. *If $(\log j)^{1/2}a_j$ is decreasing and*

$$a_j \le \gamma a_{2j} \quad \text{for some } \gamma\ge 1, \tag{5.4}$$

then

$$\max\{c_k(T), e_k(\operatorname{aco}(T))\} \le ck^{-1/2}(\log k)^{1/2}a_k. \tag{5.5}$$

3. *If for $\alpha>0$ and $\beta\in\mathbb{R}$,*

$$a_1=a_2=1,\quad a_j=\min(1,(\log j)^{-\alpha}(\log\log j)^{\beta}),\quad j\ge 3, \tag{5.6}$$

then the estimates (5.3) and (5.5) are optimal.

Proof. We start with the proof of (5.3). Let

$$b_j := (\log j)^{1/2}a_j, \tag{5.7}$$

which by assumption is increasing. For $l\in\mathbb{N}$, by definition of $\sigma(l)$ we have

$$a_{\sigma(l)+1} < q^{-l} \le a_{\sigma(l)} = (\log\sigma(l))^{-1/2}b_{\sigma(l)}, \tag{5.8}$$

which implies, by Proposition 5.1,

$$\Theta^N(T) \le c\sup_{1\le l\le N} q^{-l}\sqrt{\log\sigma(l)} \le c\sup_{1\le l\le N} b_{\sigma(l)} = cb_{\sigma(N)} \le cb_{\sigma(N)+1} = c(\log(\sigma(N)+1))^{1/2}a_{\sigma(N)+1} \le c(\log\sigma(N))^{1/2}q^{-N}.$$

Hence, in view of Theorem 4.1 it follows that

$$\sqrt{k}\,c_k(T) \le cq^{-N}[(\log\sigma(N))^{1/2}+\sqrt{k}] \tag{5.9}$$

for all $k,N\in\mathbb{N}$. Now for given $k\in\mathbb{N}$ we choose N such that

$$q^{-N-1}\le a_{2^k} < q^{-N},$$

which implies $\sigma(N)\le 2^k$. Combining these estimates with (5.9) shows (5.3) for the Gelfand width. Of course, the same arguments also imply the corresponding inequality for the entropy numbers of $\operatorname{aco}(T)$.

To verify (5.5), define, for $k\in\mathbb{N}$,

$$M(k) := \max\{M\ge 1 : q^{-M}\ge \varepsilon_k(T)\},$$
$$m(k) := \max\{M\ge 1 : q^{-M}\ge a_k\}.$$

Then $M(k)\ge m(k)$ since $\varepsilon_k(T)\le a_k$. Hence, by Theorem 4.2 and Proposition 5.2,

$$\sqrt{k}\,c_{2k}(T) \le c\Delta^{M(k)}(T) \le c\sup_{l\ge M(k)} q^{-l}\sqrt{\log\sigma(l)} \le c\sup_{l\ge m(k)} q^{-l}\sqrt{\log\sigma(l)}. \tag{5.10}$$

Define the b_j's again by (5.7), but this time they are assumed to be decreasing. From (5.10) we have

$$(5.11)\quad \sqrt{k}\, c_{2k}(T) \le c \sup_{l\ge m(k)} b_{\sigma(l)} = cb_{\sigma(m(k))} = ca_{\sigma(m(k))}(\log\sigma(m(k)))^{1/2} \le ca_{\sigma(m(k))+1}(\log\sigma(m(k)))^{1/2},$$

where we used (5.4) in the last line. By definition of $\sigma(m(k))$ and of $m(k)$ we easily get

$$(5.12)\qquad a_{\sigma(m(k))+1} < q^{-m(k)} = q\cdot q^{-m(k)-1} < qa_k.$$

On the other hand, $a_{k+1} < a_k \le q^{-m(k)}$, which implies $\sigma(m(k)) \le k$. Summing up, estimates (5.11) and (5.12) combined with $\sigma(m(k)) \le k$ yield

$$c_{2k}(T) \le ck^{-1/2}(\log k)^{1/2}a_k.$$

This proves (5.5) for Gelfand widths by (5.4). The corresponding estimate for entropy numbers is then a consequence of (5.4) and Proposition 2.1 or Theorem 4.3, respectively.

To verify that (5.3) and (5.5) are optimal, define the a_j's by (5.6) and choose t_j's in H orthogonal with $\|t_j\| = a_j$. Using the results of C. Schütt [15] and A. Garnaev and E. Gluskin [11] as in [6], we may estimate $e_k(\mathrm{aco}(T))$ as well as $c_k(T)$ from below by

$$c\cdot\begin{cases} k^{-\alpha}(\log k)^{\beta}, & \alpha < 1/2,\\ k^{-1/2}(\log k)^{-(2\alpha-1)/2}(\log\log k)^{\beta}, & \alpha > 1/2,\\ k^{-1/2}(\log k)^{\beta}, & \alpha = 1/2,\ \beta\ge 0,\\ k^{-1/2}(\log\log k)^{\beta}, & \alpha = 1/2,\ \beta < 0,\end{cases}$$

i.e. (5.3) and (5.5) cannot be improved for those a_j's. This completes the proof of the theorem.

REMARK. The preceding theorem partially answers a question raised in [1]. There the authors asked for a direct estimate of the entropy of $\mathrm{co}(T)$, $T = \{t_1, t_2, \ldots\}$, $\|t_j\| \le a_j$, by a function of the a_j's. Estimates (5.3) and (5.5) are of this form. However, (5.5) does not lead to sharp estimates in the fast decay case, i.e. for polynomial a_j's. Here (5.5) gives an extra square root of a log-term and the results of [5] and [16] are better in this case.

Because of its importance, let us state a special case of Theorem 5.1 separately.

COROLLARY 5.1. *Let T be as before and suppose $a_j = (\log j)^{-1/2}J(\log j)$ for some slowly varying function J. Then*

$$e_k(\mathrm{aco}(T)) \le \begin{cases} ck^{-1/2}J(k) & \text{for } J \text{ increasing},\\ ck^{-1/2}J(\log k) & \text{for } J \text{ decreasing}.\end{cases}$$

References

[1] K. Ball and A. Pajor, *The entropy of convex bodies with "few" extreme points*, in: London Math. Soc. Lecture Note Ser. 158, Cambridge Univ. Press, 1990, 25–32.
[2] B. Bühler, W. V. Li and W. Linde, *Localization of majorizing measures*, in: Asymptotic Methods in Probability and Statistics with Applications, Birkhäuser, to appear.
[3] B. Carl, *Entropy numbers, s-numbers, and eigenvalue problems*, J. Funct. Anal. 41 (1981), 290–306.
[4] —, *Metric entropy of convex hulls in Hilbert spaces*, Bull. London Math. Soc. 29 (1997), 452–458.
[5] B. Carl and D. E. Edmunds, *Entropy of $C(X)$-valued operators and diverse applications*, preprint, 1998.
[6] B. Carl, I. Kyrezi and A. Pajor, *Metric entropy of convex hulls in Banach spaces*, J. London Math. Soc., to appear.
[7] B. Carl and I. Stephani, *Entropy, Compactness and Approximation of Operators*, Cambridge Univ. Press, Cambridge, 1990.
[8] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Funct. Anal. 1 (1967), 290–330.
[9] —, *Universal Donsker classes and metric entropy*, Ann. Probab. 15 (1987), 1306–1326.
[10] X. Fernique, *Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens*, Les Publications CRM, Montréal, 1997.
[11] A. Garnaev and E. Gluskin, *On diameters of the Euclidean ball*, Dokl. Akad. Nauk SSSR 277 (1984), 1048–1052 (in Russian).
[12] M. Ledoux, *Isoperimetry and Gaussian analysis*, in: Lectures on Probability Theory and Statistics, Lecture Notes in Math. 1648, Springer, 1996, 165–294.
[13] M. Ledoux and M. Talagrand, *Probability in a Banach Space*, Springer, 1991.
[14] G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge Univ. Press, Cambridge, 1989.
[15] C. Schütt, *Entropy numbers of diagonal operators between symmetric Banach spaces*, J. Approx. Theory 40 (1984), 121–128.
[16] I. Steinwart, *Entropy of $C(K)$-valued operators*, J. Approx. Theory, to appear.
[17] M. Talagrand, *Regularity of Gaussian processes*, Acta Math. 159 (1987), 99–149.
[18] —, *Majorizing measures: The generic chaining*, Ann. Probab. 24 (1996), 1049–1103.

Department of Mathematics
University of Delaware
Newark, DE 19711, U.S.A.
E-mail: wli@math.udel.edu

Institut für Stochastik
FSU Jena
Ernst-Abbe-Platz 1–4
07743 Jena, Germany
E-mail: lindew@minet.uni-jena.de

Received August 10, 1998 (4160)