Entropy numbers of convex hulls and an application to learning algorithms

By

I. Steinwart

Abstract. Given a positive sequence $a = (a_n) \in \ell_{p,q}$, for $0 < p < 2$ and $0 < q \leq \infty$, and a finite set $A = \{x_1, \ldots, x_m\} \subset \ell_2$ with $|x_i| \leq a$ for all $i = 1, \ldots, m$ we prove

$$\|e_n(aco A)\|_{p,q} \leq c_{p,q} \sqrt{\log(m+1)} \|a\|_{p,q},$$

where $e_n(aco A)$ is the $n$th dyadic entropy number of the absolutely convex hull $aco A$ of $A$ and $c_{p,q} > 0$ is a suitable constant only depending on $p$ and $q$. Moreover we show that this is asymptotically optimal in $m$ for the most interesting case $q = \infty$.

As an application we give an upper bound for the so-called growth function which is of special interest in the theory of learning algorithms.

1. Introduction and results. Let $A$ be a bounded subset of a Banach space $E$, then the $n$th dyadic entropy number of $A$ is defined by

$$e_n(A) := \inf \left\{ \varepsilon > 0 : \exists x_1, \ldots, x_{2^n-1} \in K, \text{ such that } A \subset \bigcup_{k=1}^{2^n-1} (x_k + \varepsilon B_E) \right\},$$

where $B_E$ denotes the closed unit ball of the Banach space $E$. For an (bounded, linear) operator $T : E \to F$ we write $e_n(T) := e_n(TB_E)$. Sequences of entropy numbers are denoted by $(e_n(A))$ and $(e_n(T))$, respectively. Given a finite set $A = \{x_1, \ldots, x_m\} \subset E$ let $T_A : \ell^m \to E$ be the linear operator which is defined by $T_Ae_i := x_i$ where $e_1, \ldots, e_m$ is the canonical basis of $\ell^m$. Here $\ell_q$, $1 \leq q \leq \infty$, is the Banach space of all $q$-summable sequences and $\ell^n_q$ is its $n$-dimensional counterpart. Denoting by $aco A$ the absolutely convex hull of $A$ we have $aco A = T_A B_{\ell^m}$ and therefore $e_n(aco A) = e_n(T_A)$. Entropy numbers of absolutely convex hulls have recently become a subject of great interest (cf. [5], [6], [17], [13], [10] and [8]). In this article we treat a similar problem which concerns absolutely convex hulls with an additional constraint. To be precise, for a positive sequence $a = (a_i)$ we define

$$B(a) := \{(x_i) \in \ell_2 : |x_i| \leq a_i \text{ for all } i \geq 1\}.$$
Denoting the diagonal operator which maps a sequence \((x_i)\) to the sequence \((a_i x_i)\) by \(D_a\) we observe that \(B(a)\) is the image of the unit ball \(B_{\ell_\infty}\) under the diagonal operator \(D_a : \ell_\infty \to \ell_2\). We are interested in estimates of the absolutely convex hull of finitely many points of \(B(a)\) where the sequence \(a\) itself belongs to a Lorentz sequence space. To recall the latter let \(x = (x_i)\) be a sequence of real numbers. By \((s_n(x))\) we denote the non-increasing rearrangement of \(x\), that is \(s_n(x) := \inf\{c \geq 0 : \text{card}\{i : |x_i| \geq c\} < n\}\). For \(0 < p < \infty\) and \(0 < q \leq \infty\) the Lorentz sequence space \(\ell_{p,q}\) is then defined by

\[
\ell_{p,q} := \{x : (n^{1/p - 1/q} s_n(x)) \in \ell_q\},
\]

which is equipped with the quasi-norm \(\|x\|_{p,q} := \|(n^{1/p - 1/q} s_n(x))\|_{\ell_q}\).

For brevity’s sake let \(\ell_{p,q}^+ := \{x \in \ell_{p,q} : x \geq 0\}\). Given two positive sequences \((x_n)\) and \((y_n)\) we write \(x_n \preceq y_n\) if there is a constant \(c > 0\) such that \(x_n \leq c y_n\) for all \(n \geq 1\). Moreover, we write \(x_n \sim y_n\) if both hold \(x_n \preceq y_n\) and \(y_n \preceq x_n\).

We begin with an easy observation which can be derived immediately from [3, Prop. 2]:

**Proposition 1.1.** Let \(0 < p < 2\), \(0 < q \leq \infty\) and \(\frac{1}{r} := \frac{1}{p} - \frac{1}{2}\). Then there exists a constant \(c_{p,q} \geq 1\) such that for all \(a \in \ell_{p,q}^+\) and all finite sets \(A := \{x_1, \ldots, x_m\} \subset B(a)\) we have

\[
\|\{e_n(a \circ B(A))\}\|_{r,q} \leq c_{p,q} \|a\|_{p,q}.
\]

The following theorem shows how the \((p, q)\)-norm of the entropy numbers can be estimated in the above situation:

**Theorem 1.2.** Let \(0 < p < 2\) and \(0 < q \leq \infty\). Then there exists a constant \(c_{p,q} \geq 1\) such that for all \(a \in \ell_{p,q}^+\) and all \(A := \{x_1, \ldots, x_m\} \subset B(a)\) we have

\[
\|\{e_n(a \circ B(A))\}\|_{p,q} \leq c_{p,q} \sqrt{\log(m + 1)} \|a\|_{p,q}.
\]

For applications in learning theory we are mainly interested in the estimate for \(q = \infty\).

**Example 1.3.** Let \(0 < p < 2\) and \(a_i := i^{-1/p}, i \geq 1\). Moreover, for \(k \in \mathbb{N}\) and \(m := 2^k\) we define

\[A_m := \{x \in \ell_2 \mid x_i = \pm a_i \text{ for } i \leq k \text{ and } x_i = 0 \text{ for } i > k\}.
\]

Clearly, we have \(A_m \subset B(a)\), \(\text{card } A_m = m\) and \(\|a\|_{p,\infty} = 1\). Furthermore, there exists a constant \(c_p > 0\) such that for all \(m \geq 1\) we have

\[
\|\{e_n(a \circ B(A_m))\}\|_{p,\infty} \geq c_p \sqrt{\log_2 m}.
\]