CHAINING, INTERPOLATION, AND CONVEXITY

RAMON VAN HANDEL

ABSTRACT. We show that classical chaining bounds on the suprema of random processes in terms of entropy numbers can be systematically improved when the underlying set is convex: the entropy numbers need not be computed for the entire set, but only for certain "thin" subsets. This phenomenon arises from the observation that real interpolation can be used as a natural chaining mechanism. Unlike the general form of Talagrand's generic chaining method, which is sharp but often difficult to use, the resulting bounds involve only entropy numbers but are nonetheless sharp in many situations in which classical entropy bounds are suboptimal. Such bounds are readily amenable to explicit computations in specific examples, and we discover some old and new geometric principles for the control of chaining functionals as special cases.

1. INTRODUCTION

A remarkable achievement of modern probability theory is the development of sharp connections between the boundedness of random processes and the geometry of the underlying index set. Perhaps the most fundamental result in this direction is the characterization of boundedness of Gaussian processes due to Talagrand.

Theorem 1.1 ([16]). *Let $(X_t)_{t \in T}$ be a centered Gaussian process and denote by $d(t, s) = (\mathbf{E}|X_t - X_s|^2)^{1/2}$ the associated natural metric on T. Then*

$$\mathbf{E}\left[\sup_{t \in T} X_t\right] \asymp \gamma_2(T) := \inf \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} d(t, T_n),$$

where the infimum is taken over all sequences of sets T_n with cardinality $|T_n| < 2^{2^n}$.

The quantity $\gamma_2(T)$ captures precisely what aspect of the geometry of the metric space (T, d) controls the suprema of Gaussian processes: it quantifies the degree to which T can be approximated by a sequence of increasingly fine nets T_n. While we quote this particular result for concreteness, the structure that is expressed by Theorem 1.1, called the generic chaining, extends far beyond the theory of Gaussian processes and has a substantial impact on various problems in probability, functional analysis, statistics, and theoretical computer science. An extensive development of this theory and its implications can be found in [16].

Theorem 1.1 provides a powerful general principle for the study of the suprema of random processes. However, when presented with any specific situation, it often proves to be remarkably difficult to control $\gamma_2(T)$ efficiently. Theorem 1.1 can only

---

give sharp results if one is able to construct a nearly optimal sequence of nets T_n, a task that is significantly complicated by the multiscale nature of $\gamma_2(T)$. The aim of this paper is to exhibit some surprisingly elementary principles that make it possible to obtain sharp control of $\gamma_2(T)$ in various interesting examples, and that shed new light on the underlying geometric phenomena.

There are essentially two general approaches that have been used to control $\gamma_2(T)$. The simplest and by far the most useful approach is obtained by bringing the supremum over $t \in T$ inside the sum in the definition of $\gamma_2(T)$. This yields

$$\gamma_2(T) \leq \sum_{n \geq 0} 2^{n/2} e_n(T),$$

where the entropy number $e_n(T)$ is defined as the smallest $\varepsilon > 0$ such that there is an ε-net in T of cardinality less than 2^{2^n}. This bound, due to Dudley [7], long predates Theorem 1.1 and has found widespread use. Its utility stems from the fact that controlling entropy numbers only requires us to approximate the set T at a single scale, for which numerous methods are available; see, e.g., [10, 8, 2]. Unfortunately, Dudley's bound can fail to be sharp even in the simplest examples, such as ellipsoids in Hilbert space. In fact, the supremum of a random process on T cannot in general be understood in terms of the entropy numbers of T: one can easily construct two such sets with comparable entropy numbers on which a Gaussian process behaves very differently [14]. It is therefore a crucial feature of Theorem 1.1 that the use of entropy numbers is replaced by a genuinely multiscale form of approximation. The construction of such a multiscale approximation in any given situation is however a highly nontrivial task.

The main approach that has been developed for the latter purpose is Talagrand's growth functional machinery [16] that forms the core of the proof of Theorem 1.1. To show that $\gamma_2(T)$ is upper bounded by the expected supremum of the Gaussian process, the proof of Theorem 1.1 constructs nets T_n by means of a greedy partitioning scheme that uses the Gaussian process itself $G(A) := \mathbf{E}[\sup_{t \in A} X_t]$ as an objective function. It turns out that the success of this proof relies on the properties of Gaussian processes only through the validity of a single "growth condition" of the functional G. If one can design another functional F that mimics this property of Gaussian processes, then the same proof also yields an upper bound on $\gamma_2(T)$ in terms of $F(T)$. An important example of such a construction is the proof that $\gamma_2(T)$ is strictly smaller than Dudley's bound when T is a q-convex body [16, §4.1]. It is generally far from obvious, however, how a functional F can be designed, and successful application of this approach requires considerable ingenuity.

In this paper, we develop a new approach that is intermediate between these two extremes. The central insight of this paper is that it is possible to improve systematically on Dudley's bound without giving up the formulation in terms of entropy numbers. Of course, as was noted above, we cannot expect to improve on Dudley's bound in a general setting in terms of the entropy numbers of T itself. Instead, we will show that when T is a convex set, the entropy numbers $e_n(T)$ in Dudley's bound can be replaced by the entropy numbers of certain "thin" subsets that can be substantially smaller than T. (The convexity assumption is not essential for our approach, but leads to a cleaner statement of the results.)

To illustrate this idea, let us begin by stating a useful form of such a result. Let $(X, \|\cdot\|)$ be a Banach space, and let $B \subset X$ be a symmetric compact convex set. We denote by $\|\cdot\|_B$ the gauge of B, and by $\|\cdot\|_B^*$ and $\|\cdot\|^*$ the dual norms on X^*.

In this setting, we will always choose the distance d in the definitions of $\gamma_2(B)$ and $e_n(B)$ to be the one generated by the norm $d(x,y) := \|x - y\|$.

Theorem 1.2. *Let $B \subset (X, \|\cdot\|)$ be a symmetric compact convex set, and define*

$$B_t := \{y \in B : \exists\, z \in X^* \text{ such that } \langle z, y \rangle = \|y\|_B, \ \|z\|_B^* \leq 1, \ \|z\|^* \leq t\}.$$

Then we have for any $a > 0$

$$\gamma_2(B) \lesssim \frac{1}{a} + \sum_{n \geq 0} 2^{n/2} e_n(B_{a2^{n/2}}).$$

The bound of Theorem 1.2 proves to be sharp in many situations in which Dudley's bound is suboptimal, and often provides a simple explanation for why this is the case. At the same time, Theorem 1.1 is typically no more difficult to apply than Dudley's bound, as the "thin" subsets $B_t \subseteq B$ that appear in this bound can be found in quite explicit form. For example, if B is a smooth symmetric convex body in $\mathbb{R}^d$, then it is a classical fact that $\nabla \|x\|_B$ is the unique norming functional for the norm $\|\cdot\|_B$ at the point x, so that we can simply write

$$B_t = \{y \in B : \|\nabla\|y\|_B\|^* \leq t\}.$$

Such expressions are readily amenable to explicit computations.

One of the nice features of Theorem 1.2 is that the phenomenon that it describes arises in a completely elementary fashion. To understand its origin, let us sketch the simple idea behind the proof. The basic challenge in controlling $\gamma_2(B)$ is to approximate the unit ball of the norm $\|\cdot\|_B$ in terms of another norm $\|\cdot\|$. It proves to be useful to connect these two norms using an idea that is inspired by real interpolation of Banach spaces [4]. To this end, define Peetre's K-functional

$$K(t,x) := \inf_y \{\|y\|_B + t\|x - y\|\} = \|\pi_t(x)\|_B + t\|x - \pi_t(x)\|,$$

where $\pi_t(x)$ is any minimizer in the definition of $K(t,x)$ (assume for simplicity that we work in a finite-dimensional Banach space to avoid trivial technicalities). It is easily seen that $\lim_{t \to \infty} K(t,x) = \|x\|_B$, $K(0,x) = 0$, and $\frac{d}{dt}K(t,x) = \|x - \pi_t(x)\|$ (the latter follows by observing that $\|x - \pi_t(x)\|$ is a supergradient of the concave function $t \mapsto K(t,x)$, so it must equal $\frac{d}{dt}K(t,x)$ a.e.; see Proposition 2.3 below.) We therefore obtain by the fundamental theorem of calculus

$$\|x\|_B = \int_0^\infty \|x - \pi_t(x)\|\, dt \asymp \sum_{n \geq 0} 2^{n/2}\|x - \pi_{2^{n/2}}(x)\|,$$

where the last step follows from a Riemann sum approximation of the integral. This leads immediately to the following observation: if we define the sets

$$B_t := \{\pi_t(x) : x \in B\},$$

then we have shown that

$$\sup_{x \in B} \sum_{n \geq 0} 2^{n/2} d(x, B_{2^{n/2}}) \lesssim 1.$$

In other words, we see that a natural chaining mechanism is in fact built into the real interpolation method: we automatically generate a multiscale approximation of B in terms of the sets B_t. In order to bound $\gamma_2(B)$, it remains to choose a finite net with the appropriate cardinality inside each of the sets B_t. (While it may not

be immediately obvious, the definition of B_t given in Theorem 1.2 is none other than the dual formulation of the definition of B_t as a set of minimizers.)

It should be clear at this point that convexity is not essential in the construction using real interpolation: convexity only enters the proof of Theorem 1.2 in order to obtain the convenient formulation of the sets B_t. In section 2, we first prove a general form of Theorem 1.2 that is applicable in any metric space; we also formulate the results for more general γ_p-functionals that appear when the generic chaining method is applied to non-Gaussian processes. We then specialize to the convex setting and derive the dual formulation of B_t. In section 3, we illustrate the power of Theorem 1.2 in a number of explicit examples. We also illustrate by means of an example that Theorem 1.2 does not always give sharp results.

Theorem 1.2 improves on Dudley's bound by replacing the entropy numbers of B by the entropy numbers of the smaller sets B_t. A rather different improvement arises when B is q-convex, for which Talagrand shows that [16, §4.1]

$$\gamma_2(B) \lesssim \left[\sum_{n \geq 0} \left(2^{n/2} e_n(B) \right)^{q/(q-1)} \right]^{(q-1)/q}.$$

This bound involves only the entropy numbers of the set B itself, and appears at first sight to be quite different in nature than Theorem 1.2. Nonetheless, we show in section 4 that this fundamental result is a direct consequence of Theorem 1.2. Roughly speaking, we will see that the q-convexity assumption forces the sets B_t to be much smaller than the original set B in the sense that $e_n(B_t) \lesssim t^{1/(q-1)} e_n(B)^{q/(q-1)}$. In fact, it turns out there is nothing particularly special about uniform convexity: Talagrand's result is a special case of a more general geometric phenomenon that will be developed in section 4. As another illustration of this phenomenon, we will show that Talagrand's bound for q-convex bodies holds verbatim for ℓ_q-balls in Banach spaces with an unconditional basis for every $1 < q < \infty$. Note that such sets are only 2-convex rather than q-convex when $1 < q < 2$, so that the behavior of ℓ_q-balls is evidently not explained by uniform convexity.

The connection between interpolation and generic chaining appears in hindsight to be entirely natural. Many generic chaining constructions (that appear in [16, 15], for example) have a flavor of interpolation, and even the multiscale notion of approximation that is intrinsic to the definition of $\gamma_2(T)$ has appeared independently in interpolation theory in the study of approximation spaces [13, 6, 12]. To the best of the author's knowledge, however, the results of this paper are the first to explicitly develop this connection. It would be interesting to understand whether broader interactions exist between these areas of probability and analysis.

2. Chaining, Interpolation, and Convexity

The aim of this section is to develop the basic connections between chaining, interpolation, and convexity that lie at heart of this paper. In section 2.1, we develop an abstract chaining principle that holds in any metric space. In section 2.2, we specialize to the convex setting and complete the proof of Theorem 1.2.

2.1. **Chaining and interpolation.** In this section, let (X, d) be any metric space. We begin by defining formally the notions of entropy numbers and Talagrand's γ_p-functionals. The case $p = 2$ arises in the context of Gaussian processes together

with the associated natural metric, cf. Theorem 1.1; however, other values of p and more general metrics can arise for other random processes [16].

Definition 2.1. For any $A \subseteq X$ and $n \geq 0$, define the entropy number

$$e_n(A) := \inf_{|\tilde{A}| < 2^{2^n}} \sup_{x \in A} d(x, \tilde{A}),$$

and define for $p > 0$ the γ_p-functional

$$\gamma_p(A) := \inf_{|\tilde{A}_n| < 2^{2^n}} \sup_{x \in A} \sum_{n \geq 0} 2^{n/p} d(x, \tilde{A}_n).$$

(The approximating sets $\tilde{A}_n \subseteq X$ are not necessarily subsets of A.)

Fix a set $A \subseteq X$ for the remainder of this section. To measure the size of A, we introduce a penalty function $f : X \to \mathbb{R}_+ \cup \{+\infty\}$ that may in principle be chosen arbitrarily. Consider the corresponding optimization problem

$$K(t, x) := \inf_{y \in X} \{ f(y) + t d(x, y) \}$$

for every $t \geq 0$ and $x \in A$. We will assume for simplicity that the infimum in this optimization problem is attained for every $t \geq 0$ and $x \in A$, and denote by $\pi_t(x)$ any choice of minimizer in the definition of $K(t, x)$. (It is a trivial exercise to extend our results to the setting where $\pi_t(x)$ is a near-minimizer, but such an extension will not be needed in the sequel.) We now define for every $t \geq 0$ the set

$$A_t := \{ \pi_t(x) : x \in A \}.$$

Remark 2.2. In the present formulation, A_t is not necessarily a subset of A. However, it is natural to choose a penalty function f such that $A = \{x : f(x) \leq 1\}$, in which case evidently $A_t \subseteq A$ (because $f(\pi_t(x)) \leq K(t, x) \leq f(x)$).

The following result lies at the heart of this paper. In the sequel, we write $a \lesssim b$ if $a \leq Cb$ for a universal constant C, and $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. We indicate explicitly when the universal constant depends on some parameter in the problem.

Proposition 2.3. *In the setting of this section, we have for every $a > 0$*

$$\gamma_p(A) \lesssim \frac{1}{a} \sup_{x \in A} f(x) + \sum_{n \geq 0} 2^{n/p} e_n(A_{a 2^{n/p}}),$$

where the universal constant depends on p only.

Proof. We can assume without loss of generality that f is uniformly bounded on A. Thus $0 \leq K(t, x) \leq f(x) < \infty$ for every $x \in A$ and $t \geq 0$. Moreover, $t \mapsto K(t, x)$ is clearly a concave function for every $x \in A$. We now use some basic facts about univariate concave functions [9, Chapter I]. First, we note that

$$K(t, x) - K(s, x) = \inf_{y \in X} \{ f(y) + t d(x, y) \} - f(\pi_s(x)) - s d(x, \pi_s(x))$$
$$\leq (t - s) d(x, \pi_s(x))$$

for all $t, s \geq 0$, so that $d(x, \pi_s(x))$ is a supergradient of $t \mapsto K(t, x)$ at $t = s$. As a bounded concave function is absolutely continuous, we obtain

$$K(T, x) = K(0, x) + \int_0^T d(x, \pi_t(x)) \, dt$$

for every $T \geq 0$ and $x \in A$. In particular, we can estimate

$$\int_0^\infty d(x, \pi_t(x)) \, dt \leq f(x)$$

for every $x \in A$. We also recall that the derivative of a concave function is nonincreasing, so that we can discretize the integral as follows:

$$f(x) \geq \int_0^a d(x, \pi_t(x)) \, dt + \sum_{n \geq 1} \int_{a2^{(n-1)/p}}^{a2^{n/p}} d(x, \pi_t(x)) \, dt$$

$$\geq (1 - 2^{-1/p}) \, a \sum_{n \geq 0} 2^{n/p} d(x, \pi_{a2^{n/p}}(x)),$$

where we used that $t \mapsto d(x, \pi_t(x))$ is nonincreasing in the last step.

It remains to discretize the minimizers $\pi_t(x)$. By the definition of entropy numbers, we can choose for every $n \geq 0$ a set $\tilde{A}_n \subseteq X$ such that $|\tilde{A}_n| < 2^{2^n}$ and

$$\sup_{x \in A_{a2^{n/p}}} d(x, \tilde{A}_n) \leq 2 e_n(A_{a2^{n/p}}).$$

We can therefore estimate

$$\gamma_p(A) \leq \sup_{x \in A} \sum_{n \geq 0} 2^{n/p} d(x, \tilde{A}_n)$$

$$\leq \sup_{x \in A} \sum_{n \geq 0} 2^{n/p} d(x, \pi_{a2^{n/p}}(x)) + \sum_{n \geq 0} 2^{n/p} \sup_{x \in A} d(\pi_{a2^{n/p}}(x), \tilde{A}_n)$$

$$\lesssim \frac{1}{a} \sup_{x \in A} f(x) + \sum_{n \geq 0} 2^{n/p} e_n(A_{a2^{n/p}}),$$

which completes the proof.   □

Remark 2.4. Suppose we replace the penalty f by an equivalent penalty $\tilde{f} \asymp f$. Then the first term in the bound of Proposition 2.3 only changes by a universal constant, but the second term might change substantially as the definition of the sets A_t is highly nonlinear. This highlights the nontrivial nature of the choice of penalty. Similarly, the bound of Theorem 1.2 could potentially give better results if we replace B by an equivalent set $c\tilde{B} \subseteq B \subseteq C\tilde{B}$. Note that the same phenomenon arises when applying the growth functional machinery of [16]: the growth condition is not preserved if we choose an equivalent functional. This appears to be an inherent difficulty that arises in the control of chaining functionals.

2.2. **Convexity.** While Proposition 2.3 provides a very general chaining principle in metric spaces, it is not immediately obvious how to apply this result in any given situation. The problem is that the sets A_t that appear in the previous section are defined implicitly as families of solutions to certain optimization problems; in the absence of a more explicit characterization, the computation of the entropy numbers $e_n(A_{a2^{n/p}})$ can be a challenging problem. To address this problem, we specialize our results from this point onwards to the case where the set of interest is convex and where the penalty function is chosen to be the associated gauge. The convexity assumption makes it possible to obtain a dual formulation of the sets of optimizers that is readily amenable to explicit computations. The advantages of this formulation will be amply illustrated in the following sections.

We now introduce the setting that will be used throughout the remainder of this paper. Let $(X, \|\cdot\|)$ be a Banach space, and let $B \subset X$ be a symmetric compact convex set. The metric d that appears in the definitions of the entropy numbers $e_n(B)$ and the functionals $\gamma_p(B)$ (cf. Definition 2.1) will always be chosen to be defined by the norm $d(x,y) := \|x - y\|$ on the underlying Banach space. The gauge (Minkowski functional) of B will be denoted $\|\cdot\|_B$, that is,

$$\|x\|_B := \inf\{s \geq 0 : x \in sB\}$$

for $x \in X$. Denote by $\|\cdot\|_B^*$ and $\|\cdot\|^*$ the associated dual gauge and norm, that is,

$$\|z\|_B^* := \sup_{\|x\|_B \leq 1} \langle z, x \rangle = \sup_{x \in B} \langle z, x \rangle, \qquad \|z\|^* := \sup_{\|x\| \leq 1} \langle z, x \rangle$$

for $z \in X^*$. The key point of this section is the following duality result, which shows that the minimizers of the K-functional in the convex setting define a form of projection onto an explicitly defined scale of subsets $B_t \subseteq B$.

Proposition 2.5. *For every $t \geq 0$, there is a map $\pi_t : B \to B$ such that:*

(i) $\pi_t(x)$ is a minimizer for Peetre's K-functional for every $x \in B$:

$$K(t,x) := \inf_{y \in X}\{\|y\|_B + t\|x - y\|\} = \|\pi_t(x)\|_B + t\|x - \pi_t(x)\|.$$

(ii) The set of minimizers

$$B_t := \{\pi_t(x) : x \in B\}$$

can be characterized as

$$B_t = \{y \in B : \exists z \in X^* \text{ such that } \langle z, y \rangle = \|y\|_B, \ \|z\|_B^* \leq 1, \ \|z\|^* \leq t\}.$$

(iii) We have $\pi_t(x) = x$ for every $x \in B_t$.

Proof. The result holds trivially for $t = 0$, so we fix $t > 0$ in the sequel.

Step 1. Let $B_K := \text{conv}(B \cup \frac{1}{t}B_\sim)$, where $B_\sim$ is the closed unit ball in $(X, \|\cdot\|)$. For completeness, we recall the proof of the elementary fact that $K(t,x) = \|x\|_{B_K}$ for every $x \in X$, where $\|\cdot\|_{B_K}$ denotes the gauge of B_K.

Suppose first that $K(t,x) < r$, so there exists $y \in X$ with $\|y\|_B + t\|x - y\| < r$. Then writing $x = \lambda x_1 + \mu x_2$ with $x_1 = y/\|y\|_B$ and $x_2 = (x - y)/t\|x - y\|$ readily implies that $\|x\|_{B_K} < r$. In the converse direction, suppose that $\|x\|_{B_K} < r$, so that $x = \lambda x_1 + \mu x_2$ for some $|\lambda| + |\mu| < r$, $x_1 \in B$, $x_2 \in \frac{1}{t}B_\sim$. Then choosing $y = \lambda x_1$ in the definition of $K(t,x)$ shows that $K(t,x) < r$.

Step 2. We now establish the existence of a minimizer in the definition of $K(t,x)$ for every $x \in X$. This is a direct consequence of the previous step and the compactness of B. Indeed, as B is compact, the set B_K is closed. Thus $K(t,x) = r$ implies $x \in rB_K$, so there exist $|\lambda| + |\mu| \leq r$ and $x_1 \in B$, $x_2 \in \frac{1}{t}B_\sim$ such that $x = \lambda x_1 + \mu x_2$. It follows that $y = \lambda x_1$ is a minimizer for $K(t,x)$, as

$$K(t,x) \leq \|\lambda x_1\|_B + t\|\mu x_2\| \leq r = K(t,x).$$

Step 3. Define the set

$$B_t' := \{y \in B : K(t,y) = \|y\|_B\}.$$

We can characterize this set by duality. Indeed, note that

$$K(t,y) = \sup\{\langle z, y \rangle : z \in X^*, \ \|z\|_B^* \leq 1, \ \|z\|^* \leq t\},$$

where we have used the polar identity $B_K^\circ = B^\circ \cap tB_\sim^\circ$. Moreover, the supremum is attained at some point $z \in X^*$ by the Hahn-Banach theorem. Therefore, if

$y \in B'_t$, then there exists $z \in X^*$ such that $\langle z, y \rangle = \|y\|_B$, $\|z\|_B^* \le 1$, and $\|z\|^* \le t$. Conversely, if $y \in B$ is such that a point z satisfying the latter properties exists, then $\|y\|_B = \langle z, y \rangle \le K(t, y) \le \|y\|_B$ so that $y \in B'_t$. Thus we have

$$B'_t = \{y \in B : \exists\, z \in X^* \text{ such that } \langle z, y \rangle = \|y\|_B,\ \|z\|_B^* \le 1,\ \|z\|^* \le t\}.$$

Step 4. Define the map $\pi_t : B \to B$ as follows. For $x \in B'_t$, we set $\pi_t(x) = x$. For $x \notin B'_t$, we choose $\pi_t(x)$ to be any minimizer in the definition of $K(t, x)$. We are going to verify that each of the claims in the statement of the Proposition hold.

Let us first note that π_t does indeed map B into itself. For $x \in B'_t$, this is true by construction. For $x \notin B'_t$, this is true because $\|\pi_t(x)\|_B \le K(t, x) \le \|x\|_B$. Moreover, note that when $x \in B'_t$, by construction $y = x = \pi_t(x)$ is a minimizer in the definition of $K(t, x)$. We have therefore established part (i).

To prove parts (ii) and (iii), it suffices to show that $B_t = B'_t$. That $B'_t \subseteq B_t$ is obvious from the fact that $\pi_t(x) = x$ for $x \in B'_t \subseteq B$. To establish the converse inclusion, we argue as follows. Fix $x \in B$, and choose $z \in X^*$ such that $K(t, x) = \langle z, x \rangle$, $\|z\|_B^* \le 1$, and $\|z\|^* \le t$. By the bipolar theorem, we can write

$$\langle z, \pi_t(x) \rangle \le \|\pi_t(x)\|_B = \langle z, \pi_t(x) \rangle + \langle z, x - \pi_t(x) \rangle - t\|x - \pi_t(x)\| \le \langle z, \pi_t(x) \rangle.$$

This implies that $\pi_t(x) \in B'_t$, and thus $B_t \subseteq B'_t$. $\square$

Remark 2.6. When B is a symmetric convex body in a finite-dimensional Banach space, the details of the proof of Proposition 2.5 simplify significantly. It is an instructive exercise to give a quick proof in this case using subdifferential calculus.

The proof of Theorem 1.2 in the introduction now follows trivially. For future reference, we formulate the analogous result for γ_p-functionals.

Corollary 2.7. *Let $B \subset (X, \|\cdot\|)$ be a symmetric compact convex set, and define*

$$B_t := \{y \in B : \exists\, z \in X^* \text{ such that } \langle z, y \rangle = \|y\|_B,\ \|z\|_B^* \le 1,\ \|z\|^* \le t\}.$$

Then we have for any $a > 0$

$$\gamma_p(B) \lesssim \frac{1}{a} + \sum_{n \ge 0} 2^{n/p} e_n(B_{a 2^{n/2}}),$$

where the universal constant depends on p only.

Proof. This is simply the combined statement of Proposition 2.3, where we choose the penalty $f(x) = \|x\|_B$ and distance $d(x, y) = \|x - y\|$, and Proposition 2.5. $\square$

We end this section by emphasizing a remark that was also made in the introduction. Recall that a symmetric convex set $B \subset X$ is called smooth if for every $x \in X$, $x \ne 0$ there is a unique $z \in X^*$ so that $\langle z, x \rangle = \|x\|_B$ and $\|z\|_B^* \le 1$, cf. [3].

Corollary 2.8. *Let B be a symmetric convex body in a finite-dimensional Banach space $(X, \|\cdot\|)$, and denote by $\partial\|y\|_B$ the subdifferential of $\|y\|_B$. Then*

$$B_t = \Big\{y \in B : \inf_{z \in \partial\|y\|_B} \|z\|^* \le t\Big\}.$$

In particular, if B is smooth, then

$$B_t = \{y \in B : \|\nabla\|y\|_B\|^* \le t\}.$$

Proof. It is a classical fact that $\partial\|y\|_B = \{z \in X^* : \langle z, y \rangle = \|y\|_B,\ \|z\|_B^* \le 1\}$, so that the result follows readily from Proposition 2.5; cf. [9, Chapter VI]. $\square$

The explicit nature of Corollary 2.8 is particularly useful in computations.

3. Examples

The aim of this section is to illustrate the utility of Theorem 1.2 in explicit computations by investigating some simple but conceptually interesting examples. As our goal is to develop insight into the phenomenon described by Theorem 1.2, we have avoided unnecessary distractions by restricting attention to situations in which existing entropy estimates can be used.

We write $\|x\|_r := [\sum_i |x_i|^r]^{1/r}$, and denote by $e_1, \ldots, e_d$ the standard basis in $\mathbb{R}^d$. Throughout this section, we work in Euclidean space $(\mathbb{R}^d, \|\cdot\|)$ where $\|\cdot\| := \|\cdot\|_2$. The concrete choice of the Euclidean norm is not important for our theory, but is made in order to enable explicit computations and is natural in the setting of Gaussian processes (as it corresponds to the canonical choice $X_t = \langle t, g \rangle$ in Theorem 1.1, where g is a standard Gaussian vector in $\mathbb{R}^d$). Some of the examples developed here will be revisited in section 4 in a much more general setting.

3.1. ℓ_q-Ellipsoids.

The classical example of a situation where Dudley's bound fails to be sharp is that of ellipsoids in Hilbert space. In this section, we will investigate the following more general situation. Given scalars $1 < q < \infty$ and $b_1 \geq b_2 \geq \cdots \geq b_d > 0$, let $B \subset \mathbb{R}^d$ be the ℓ_q-ellipsoid whose gauge is given by

$$\|x\|_B = \left[\sum_{i=1}^d \left(\frac{|x_i|}{b_i} \right)^q \right]^{1/q}.$$

We will show that Theorem 1.2 yields the following optimal bound.

Proposition 3.1. *In the setting of this section, we have*

$$\gamma_2(B) \lesssim \left(\sum_{i=1}^d b_i^{q/(q-1)} \right)^{(q-1)/q},$$

where the universal constant depends on q only.

Of course, this result can easily be obtained from Theorem 1.1, but our aim is to provide a geometric proof that explains why the result is true.

In order to apply either Dudley's bound or Theorem 1.2, we will require suitable estimates on the entropy numbers of ℓ_q-ellipsoids. The behavior of these entropy numbers is investigated in detail in a classic paper by Carl [5] (in the special case of ℓ_2-ellipsoids a much more elementary approach can be found in [16, §2.5]). For future reference, we record a more general form of the main result of Carl than is presently needed. The following can be read off from the proof of [5, Theorem 2]. (While the result of Carl is formulated only for $r \geq 1$, the proof extends directly to the case $0 < r < 1$ if we replace [5, Theorem 1] by [8, Proposition 3.2.2].)

Lemma 3.2 ([5]). *Given $0 < r < \infty$, $1/s > (1/2 - 1/r)_+$, $0 < u < \infty$, and scalars $c_1 \geq c_2 \geq \cdots \geq c_d > 0$, the ℓ_r-ellipsoid $C = \{x \in \mathbb{R}^d : \|(x_i/c_i)\|_r \leq 1\}$ satisfies*

$$\sum_{n \geq 0} \left(2^{n(1/s+1/r-1/2)} e_n(C) \right)^u \asymp \sum_{k=1}^d (k^{1/s-1/u} c_k)^u$$

where the universal constant depends on r, s, u only.

Applying this result with $r = q$, $1/s = 1 - 1/q$, and $u = 1$ yields

$$\sum_{n \geq 0} 2^{n/2} e_n(B) \asymp \sum_{k=1}^{d} k^{-1/q} b_k.$$

We therefore see immediately that Dudley's bound is suboptimal for ℓ_q-ellipsoids: Dudley's bound is much larger than $\gamma_2(B)$, say, when $b_k = k^{-(q-1)/q}(\log k)^{-1}$.

To obtain a sharp bound, we will apply Theorem 1.2. The crux of the matter is to control the sets B_t. In the present setting, this is exceedingly simple and gives a vivid illustration of where the improvement over Dudley's bound comes from.

Proof of Proposition 3.1. Note that B is a smooth convex body with

$$\frac{\partial \|y\|_B}{\partial y_k} = \frac{1}{b_k^q} \frac{|y_k|^{q-1}}{\|y\|_B^{q-1}} \operatorname{sign}(y_k).$$

Thus Corollary 2.8 gives

$$B_t = \{y \in B : \|y\|_C \leq t^{1/(q-1)} \|y\|_B\} \subseteq t^{1/(q-1)} C,$$

where

$$\|y\|_C = \left[\sum_{i=1}^{d} \left(\frac{|y_i|}{b_i^{q/(q-1)}} \right)^{2q-2} \right]^{1/(2q-2)}.$$

Substituting $B_t \subseteq t^{1/(q-1)} C$ into Theorem 1.2 and optimizing over $a > 0$ yields

$$\gamma_2(B) \lesssim \left(\sum_{n \geq 0} 2^{nq/(2q-2)} e_n(C) \right)^{(q-1)/q}.$$

The conclusion follows by applying Lemma 3.2 with $r = 2q - 2$ and $s = u = 1$. $\square$

The key point of the proof of Proposition 3.1 is that each subset B_t of the ℓ_q-ellipsoid B is contained in a dilation of the much "thinner" ℓ_{2q-2}-ellipsoid C: the lengths of the semiaxes of C have been raised to the power $q/(q-1)$ as compared to those of B. This is precisely why we obtain the correct powers of b_i inside the sum in Proposition 3.1. The author sees no obvious way to explain this miracle other than that it drops out of the trivial explicit computation performed above. However, a deeper understanding of the geometry of the sets B_t for ℓ_q-ellipsoids will be obtained in a much more general setting in section 4.

Remark 3.3. There exist two previous geometric proofs of Proposition 3.1 for special values of q. The first, in [11, §15.6], gives a delicate manual construction of an equivalent formulation of $\gamma_2(B)$ for $q = 2$. The second, in [16, §4.1], deduces the result for $2 \leq q < \infty$ from a more general bound for uniformly convex bodies that is proved using the growth functional machinery. We will revisit the latter idea in section 4, where we will also see that uniform convexity fails to explain the behavior of ℓ_q-ellipsoids for $1 < q < 2$. That we have obtained a sharp bound for every value of q with the same proof therefore hides the fact that ℓ_q-ellipsoids can have a very different geometry for different values of q.

Remark 3.4. The universal constant in Proposition 3.1 must necessarily depend on q: if this were not the case, then we would obtain $\gamma_2(B) \lesssim b_1$ in the limit $q \downarrow 1$ which is easily seen to be false by Theorem 1.1. Unfortunately, the entropy estimates provided by Lemma 3.2 are not sufficiently accurate to recover the correct

behavior as $q \downarrow 1$. This is not a deficiency of Theorem 1.2, however: the case $q = 1$ is of particular interest in its own right and will be investigated in the next section.

3.2. **Octahedra.** In this section, we investigate the limiting case $q = 1$ of the example developed in the previous section. That is, given scalars $b_1 \geq b_2 \geq \cdots \geq b_d > 0$, we investigate the octahedron $B \subset \mathbb{R}^d$ defined by

$$B = \text{absconv}\{b_i e_i : i = 1, \ldots, d\}.$$

It is not difficult to show that Dudley's bound is suboptimal in this setting [16, Exercise 2.2.15]. We will show that Theorem 1.2 yields the following optimal bound.

Proposition 3.5. *In the setting of this section, we have*

$$\gamma_2(B) \lesssim \Sigma := \max_{i \leq d} b_i \sqrt{\log(i+1)}.$$

Of course, this result could easily be obtained from Theorem 1.1, and a rather difficult geometric proof using growth functionals can be found in [15, §8]. However, the point for our purposes is that this result follows in a completely elementary fashion from Theorem 1.2. To apply the latter, let us first identify the sets B_t.

Lemma 3.6. *For any $t \geq 0$, we have*

$$B_t = \left\{ y \in B : \sum_{i=1}^d \frac{\mathbf{1}_{y_i \neq 0}}{b_i^2} \leq t^2 \right\}.$$

Proof. While $\|\cdot\|_B$ is not smooth, we can easily compute its subdifferential:

$$\partial \|y\|_B = \{z \in \mathbb{R}^d : z_i = \text{sign}(y_i)/b_i \text{ if } y_i \neq 0, \ |z_i| \leq 1/b_i \text{ if } y_i = 0\}.$$

We therefore obtain

$$\inf_{z \in \partial \|y\|_B} \|z\|^2 = \sum_{i=1}^d \frac{\mathbf{1}_{y_i \neq 0}}{b_i^2},$$

and the result follows from Corollary 2.8.   □

Lemma 3.6 shows that the sets B_t are very thin indeed: they consist of sparse vectors. Controlling the entropy numbers of such sets is an easy exercise; for each fixed sparsity pattern we can discretize using a standard volumetric argument, while counting the number of sparsity patterns is a matter of simple combinatorics.

Lemma 3.7. *There is a universal constant $c > 0$ such that for all $n \geq 0$*

$$e_n(B_{c2^{n/2}/\Sigma}) \lesssim 2^{-n} b_1.$$

Proof. Fix $n \geq 0$. As $1/b_i^2 \geq \log(i+1)/\Sigma^2$ by definition, we have

$$B_t \subseteq C_t := \left\{ y \in B : \sum_{i=1}^d \log(i+1)\, \mathbf{1}_{y_i \neq 0} \leq \Sigma^2 t^2 \right\}.$$

It suffices to control the entropy numbers of the larger set $C_{c2^{n/2}/\Sigma}$.

Let us begin with some counting. Denote by $\mathcal{I}$ the family of all admissible sparsity patterns of $y \in C_{c2^{n/2}/\Sigma}$, that is, $\mathcal{I}$ is the family of all $I \subseteq [d]$ such that

$$\sum_{i \in I} \log(i+1) \leq c^2 2^n.$$

Denote by $\mathcal{I}_k \subseteq \mathcal{I}$ the family of all $I \in \mathcal{I}$ with cardinality $|I| = k$. Let us bound the number of such sets. Setting $c := \sqrt{\log 2}/2$, we can estimate

$$|\mathcal{I}_k| = \sum_{|I|=k} \mathbf{1}_{I \in \mathcal{I}} = \sum_{|I|=k} \mathbf{1}_{\prod_{i \in I}(i+1)^2 \leq 2^{2^{n-1}}} \leq 2^{2^{n-1}} \sum_{|I|=k} \prod_{i \in I} \frac{1}{(i+1)^2}.$$

The right-hand side can be bounded as follows:

$$\sum_{|I|=k} \prod_{i \in I} \frac{1}{(i+1)^2} = \sum_{1 \leq \ell_1 < \ell_2 < \cdots < \ell_k \leq d} \prod_{i=1}^{k} \frac{1}{(\ell_i+1)^2} \leq \prod_{i=1}^{k} \sum_{\ell \geq i} \frac{1}{(\ell+1)^2} < \frac{1}{k!},$$

where we have used that

$$\sum_{\ell \geq i} \frac{1}{(\ell+1)^2} < \sum_{\ell \geq i} \int_{\ell}^{\ell+1} \frac{1}{x^2}\,dx = \int_{i}^{\infty} \frac{1}{x^2}\,dx = \frac{1}{i}.$$

We have therefore shown that $|\mathcal{I}_k| < 2^{2^{n-1}}/k!$.

Let $\varepsilon \leq b_1$ be a constant to be chosen later on. For every $I \in \mathcal{I}$, choose a minimal ε-net T_I for the Euclidean ball in $\mathbb{R}^I$ with radius b_1, and denote by T the union of all these sets T_I. Evidently T is a ε-net for $C_{c2^{n/2}/\Sigma}$. Let us estimate its cardinality. A standard volumetric argument yields [2, Corollary 4.1.15]

$$|T_I| \leq \left(\frac{3b_1}{\varepsilon}\right)^{|I|}.$$

We can therefore estimate

$$|T| \leq \sum_{k=0}^{d} \left(\frac{3b_1}{\varepsilon}\right)^{k} |\mathcal{I}_k| < 2^{2^{n-1}} e^{3b_1/\varepsilon}.$$

If we choose $\varepsilon = (6/\log 2)\,2^{-n}b_1$, we find that $|T| < 2^{2^n}$ which establishes the claim whenever $2^n \geq 6/\log 2$ (as we assumed that $\varepsilon \leq b_1$ in the volumetric estimate). For $2^n < 6/\log 2$, simply note the trivial bound $e_n(C_{c2^{n/2}/\Sigma}) \leq \mathrm{diam}(B) \leq 2b_1$. $\qquad\square$

With this entropy estimate in hand, the proof of Proposition 3.5 is an immediate consequence of Lemma 3.7 and Theorem 1.2 with $a = c/\Sigma$.

3.3. **A counterexample.** The aim of this section is to show that Theorem 1.2 does not always give sharp results. As the example that we will discuss is a conceptually important one, let us briefly consider this example in a broader context.

A remarkable consequence of Theorem 1.1 is that $\gamma_2(\mathrm{conv}(T)) \asymp \gamma_2(T)$ for any (non-convex) subset $T \subseteq \mathbb{R}^d$ of Euclidean space: as the supremum of a linear function over a convex set is attained at an extreme point, Theorem 1.1 yields

$$\gamma_2(T) \asymp \mathbf{E}\left[\sup_{x \in T} \langle x, g \rangle\right] = \mathbf{E}\left[\sup_{x \in \mathrm{conv}(T)} \langle x, g \rangle\right] \asymp \gamma_2(\mathrm{conv}(T))$$

(here g denotes a standard Gaussian vector in $\mathbb{R}^d$). It is a long-standing open problem to understand the geometric mechanism behind this fundamental fact; cf. [16, §2.4]. By using a known device [16, Theorem 2.4.18], one can reduce this problem to the following special case: it suffices to give a geometric proof of the fact that for any $x_1, \ldots, x_n \in \mathbb{R}^d$ such that $\|x_1\| \geq \|x_2\| \geq \cdots \geq \|x_n\| > 0$, we have

$$\gamma_2(B) \lesssim \max_{i \leq n} \|x_i\| \sqrt{\log(i+1)}, \qquad B = \mathrm{absconv}\{x_i : i = 1, \ldots, n\}.$$

We solved this problem in the previous section under the additional assumption that the vectors x_i are orthogonal. It is not known, however, how this conclusion can be established in the absence of the orthogonality assumption. The results of this paper originated in an attempt by the author to understand this issue. We will presently illustrate that Theorem 1.2 does not directly resolve this problem.

The example that we will consider is defined as follows. Fix $0 < \varepsilon < 1$ and let $u = d^{-1/2}\mathbf{1}$, where $\mathbf{1}$ is the vector of ones (note that $\|u\| = 1$). We consider the set

$$B = \mathrm{absconv}\{x_i : i = 1, \ldots, d\}, \qquad x_i = e_i + \varepsilon u.$$

This is a small perturbation of the example in the previous section where all vertices of the simplex have been shifted along the diagonal. One can show as in [16, Exercise 2.2.15] that $\gamma_2(B) \asymp \sqrt{\log d}$, while Dudley's bound is of order $(\log d)^{3/2}$.

We claim that Theorem 1.2 does not improve on Dudley's bound in the present setting: the sets B_t are not sufficiently small to gain any improvement. This unfortunate conclusion is contained in the following lemma.

Lemma 3.8. *We have $B_t \supseteq \mathrm{conv}\{x_i : i = 1, \ldots, d\}$ for all $t \geq 1/\varepsilon$.*

Proof. Let $V = \sum_{i=1}^{d} x_i \otimes e_i$ be the square matrix whose columns are the vectors x_i. Note that V is invertible, and we have $\|x\|_B = \|V^{-1}x\|_1$. Therefore

$$\partial\|x\|_B = (V^*)^{-1}\partial\|V^{-1}x\|_1 \ni (V^*)^{-1}\mathrm{sign}(V^{-1}x),$$

where $\mathrm{sign}(z)$ operates entrywise on a vector z and we set $\mathrm{sign}(0) := 1$. In particular,

$$B_t \supseteq \{x \in B : \mathrm{sign}(V^{-1}x) \in tV^*B_\sim\}$$

by Corollary 2.8, where $B_\sim$ denotes the Euclidean unit ball in $\mathbb{R}^d$.

Now note that if $x \in \mathrm{conv}\{x_i : i = 1, \ldots, d\}$, then $V^{-1}x$ has nonnegative entries and thus $\mathrm{sign}(V^{-1}x) = \mathbf{1}$. It therefore suffices to show that $\mathbf{1} \in tV^*B_\sim$ whenever $t \geq 1/\varepsilon$. But this is a simple consequence of the definition of V, as

$$tV^*v = \mathbf{1} \quad \text{for} \quad v = \frac{u}{t(\varepsilon + d^{-1/2})}$$

and clearly $\|v\| \leq 1$ when $t \geq 1/\varepsilon$. This completes the proof. $\qquad\square$

Let Δ^{d-1} be the standard simplex in $\mathbb{R}^d$. Lemma 3.8 shows that $B_t \supseteq \Delta^{d-1} + \varepsilon u$ whenever $t \geq 1/\varepsilon$. Setting $n_{a,\varepsilon} = (2\log_2(1/a\varepsilon))_+$, we can estimate

$$\sum_{n\geq 0} 2^{n/2}e_n(B_{a2^{n/2}}) \geq \sum_{n\geq n_{a,\varepsilon}} 2^{n/2}e_n(\Delta^{d-1}) \gtrsim (\log d)^{3/2} - Cn_{a,\varepsilon}\sqrt{\log d}$$

for some constant $C > 0$, where we used that $e_n(\Delta^{d-1}) \gtrsim 2^{-n/2}\sqrt{\log d}$ for $n \lesssim \log d$ [16, Exercise 2.2.15]. We have therefore shown that Theorem 1.2 does not improve on Dudley's bound in this example unless ε is polynomially small in d.

Remark 3.9. Of course, the example described in this section is sufficiently simple that we can make some manual adjustments to obtain a sharp geometric construction. Indeed, we clearly have $B \subset B_1 + B_2$ where B_1 denotes the ℓ_1-ball in $\mathbb{R}^d$ and $B_2 = \{\alpha u : |\alpha| \leq \varepsilon\}$ is one-dimensional. Theorem 1.2 gives a sharp generic chaining construction for B_1, while a trivial discretization of α suffices to control B_2. We can then glue together the generic chaining constructions for B_1 and B_2 by summing the corresponding nets. It is not clear, however, how one could construct such a decomposition in the general setting described at the beginning of this section.

4. Geometry and Entropy Contraction

In the previous section, we illustrated the utility of Theorem 1.2 in specific examples. The computations hinge, however, on a sufficiently explicit description of the sets B_t, which may not always be available in more general situations. For example, if we consider the examples of the previous section under general norms, it may be nontrivial to control the sets B_t directly. It is therefore of interest to develop more systematic methods to control the geometry of the sets B_t.

As a prototype of what one might hope for, let us reconsider the setting of ℓ_q-ellipsoids in Hilbert space. Theorem 1.2 bounds $\gamma_2(B)$ in terms of the entropy numbers of the sets B_t, which we computed explicitly in section 3.1. However, Lemma 3.2 suggests that the correct behavior of $\gamma_2(B)$ in this example can also be expressed in terms of the entropy numbers of B itself: we easily verify that

$$\gamma_2(B) \asymp \left[\sum_{n \geq 0} \left(2^{n/2} e_n(B) \right)^{q/(q-1)} \right]^{(q-1)/q}.$$

The appearance of such a bound is not a coincidence. Talagrand has shown that an upper bound of this form holds for any q-convex set B [16, §4.1]: as ℓ_q-ellipsoids are $\max(2, q)$-convex, this provides an alternative explanation for the behavior of ℓ_q-ellipsoids in the case $2 \leq q < \infty$. One of the insights to be developed in this section is that this fundamental property of q-convex sets is fully explained by Theorem 1.2. Roughly speaking, we will show that the q-convexity assumption forces the sets B_t to be much smaller than B itself in the sense that $e_n(B_t) \lesssim t^{1/(q-1)} e_n(B)^{q/(q-1)}$, from which the above bound is easily deduced. More generally, this phenomenon suggests that the chaining principle for general convex sets given by Theorem 1.2 can be significantly simplified in the presence of additional geometric structure.

It turns out that there is nothing special about q-convexity per se, but that the entropy contraction phenomenon illustrated above arises from a much more general geometric mechanism. We develop a general formulation of this idea in section 4.1. We then demonstrate how the requisite structure arises in two distinct settings: the case of q-convex sets is developed in section 4.2, while the case of ℓ_q-balls in Banach spaces with an unconditional basis is developed in section 4.3.

4.1. **A geometric principle.** Let $(X, \|\cdot\|)$ be a Banach space and let $B \subset X$ be a symmetric compact convex set. The sets B_t are defined as in Theorem 1.2. The following geometric principle is the main result of this section.

Theorem 4.1. *Let $q > 1$ and $K > 0$ be given constants, and suppose that*

$$\|y - z\|_B^q \leq Kt \|y - z\| \quad \text{for every } y, z \in B_t, \ t \geq 0.$$

Then

$$\gamma_p(B) \lesssim \left[\sum_{n \geq 0} \left(2^{n/p} e_n(B) \right)^{q/(q-1)} \right]^{(q-1)/q},$$

where the universal constant depends on p, q, and K only.

Like Theorem 1.2, the message of Theorem 4.1 is that the behavior of $\gamma_p(B)$ is strictly better than would be expected from Dudley's bound. Unlike Theorem 1.2, however, the presence of additional geometric structure allows us to bound $\gamma_p(B)$ only in terms of the entropy numbers of B itself. This bound could therefore be

applied even without an explicit description of B_t. Of course, there is no free lunch: the assumption of Theorem 4.1 requires us to understand the metric structure of the sets B_t. Fortunately, we will see in the sequel that there are interesting situations in which this can be accomplished without explicitly computing the sets B_t.

Remark 4.2. Before we turn to the proof of Theorem 4.1, it is instructive to consider the significance of the geometric assumption of Theorem 4.1. Observe that we always have, regardless of any assumptions, the following simple fact:

$$\|y\|_B \le t\|y\| \quad \text{for every } y \in B_t, \ t \ge 0.$$

Indeed, if $z \in X^*$ is as in the definition of B_t, then

$$\|y\|_B = \langle z, y \rangle \le \|z\|^* \|y\| \le t\|y\|.$$

We therefore see that by construction, an element $y \in B_t$ with small norm must be contained in a small dilation $y \in t\|y\|B$ of the original set B. The assumption of Theorem 4.1 asks that a weaker form of this property hold not only for norms, but also for distances: that is, if $y, z \in B_t$, then $y - z \in (Kt\|y - z\|)^{1/q}B$. This does not follow automatically from the corresponding property for norms, as it is typically not true that $B_t - B_t \subseteq cB_{ct}$ for some constant c. Nonetheless, this intuition proves to be useful as it will help us identify how the requisite geometric structure arises.

The main idea behind the proof of Theorem 4.1 is the following observation.

Lemma 4.3. *Suppose that the assumption of Theorem 4.1 holds. Then*

$$e_{n+1}(B_t) \le (Kt\, e_n(B_t))^{1/q} e_n(B) \quad \text{for every } n \ge 0, \ t \ge 0.$$

Proof. Fix $\varepsilon > 0$. By the definition of entropy numbers, we can cover B_t by less than 2^{2^n} balls of radius $(1+\varepsilon)e_n(B_t)$. By our assumption, each of these balls (intersected with B_t) is contained in a translate of sB with $s \le (1+\varepsilon)^{1/q}(Kt\, e_n(B_t))^{1/q}$. Therefore, each of these balls can be further covered by less than 2^{2^n} balls of radius $(1+\varepsilon)s\, e_n(B)$. We have now covered B_t by less than $2^{2^n} \cdot 2^{2^n} = 2^{2^{n+1}}$ balls of radius $\le (1+\varepsilon)^{1+1/q}(Kt\, e_n(B_t))^{1/q}e_n(B)$. Letting $\varepsilon \downarrow 0$ completes the proof. $\qquad\square$

An annoying feature of Lemma 4.3 is that the entropy number on the left-hand side is $e_{n+1}(B_t)$ rather than $e_n(B_t)$. If it were the case that $e_n(B_t) \lesssim e_{n+1}(B_t)$ (that is, if we knew *a priori* that the entropy numbers do not decay too quickly), then we could simplify the conclusion of Lemma 4.3 to

$$e_n(B_t) \lesssim t^{1/(q-1)}e_n(B)^{q/(q-1)}.$$

This expression quantifies in the present setting in what sense the sets B_t are much smaller than the original set B. From this expression, it would be easy to conclude the result of Theorem 4.1: substituting the above bound into Theorem 1.2 yields

$$\gamma_p(B) \lesssim \frac{1}{a} + a^{1/(q-1)} \sum_{n \ge 0} \left(2^{n/p}e_n(B)\right)^{q/(q-1)},$$

and the conclusion of Theorem 4.1 would follow by optimizing over $a > 0$. The main technical issue in the proof of Theorem 4.1 is to show that its conclusion remains valid even when the regularity assumption $e_n(B_t) \lesssim e_{n+1}(B_t)$ does not hold, which we do by means of a routine dyadic regularization argument.

Proof of Theorem 4.1. Fix a constant $\lambda > 0$ to be chosen at a later stage. For any set C, we introduce the regularized entropy numbers $d_n(C) \geq e_n(C)$ as

$$d_n(C) := \max_{0 \leq k \leq n} 2^{\lambda(k-n)} e_k(C).$$

Using Lemma 4.3, we estimate

$$d_n(B_t) \leq \max_{0 \leq k \leq n+1} 2^{\lambda(k-n)} e_k(B_t)$$

$$\leq 2^{-\lambda n} \mathrm{diam}(B) + 2^\lambda \max_{0 \leq k \leq n} 2^{\lambda(k-n)} e_{k+1}(B_t)$$

$$\lesssim 2^{-\lambda n} \mathrm{diam}(B) + 2^\lambda t^{1/q} \max_{0 \leq k \leq n} 2^{\lambda(k-n)} e_k(B_t)^{1/q} e_k(B)$$

$$\leq 2^{-\lambda n} \mathrm{diam}(B) + 2^\lambda t^{1/q} d_n(B_t)^{1/q} \max_{0 \leq k \leq n} 2^{\lambda(k-n)(q-1)/q} e_k(B).$$

Therefore, using $a^{1/q} b^{(q-1)/q} \leq a/q + b(q-1)/q$, we obtain

$$d_n(B_t) \lesssim 2^{-\lambda n} \mathrm{diam}(B) + 2^{\lambda q/(q-1)} t^{1/(q-1)} \max_{0 \leq k \leq n} 2^{\lambda(k-n)} e_k(B)^{q/(q-1)}.$$

In particular, we can crudely bound

$$\sum_{n \geq 0} 2^{n/p} e_n(B_{a2^{n/p}}) \lesssim \mathrm{diam}(B) \sum_{n \geq 0} 2^{n/p} 2^{-\lambda n} +$$

$$a^{1/(q-1)} 2^{\lambda q/(q-1)} \sum_{n \geq 0} 2^{nq/(q-1)p} 2^{-\lambda n} \sum_{0 \leq k \leq n} 2^{\lambda k} e_k(B)^{q/(q-1)}.$$

In order for the sums to converge we must choose $\lambda > q/(q-1)p$, so we fix for concreteness $\lambda = 2q/(q-1)p$ (the precise value of λ does not matter). This yields

$$\sum_{n \geq 0} 2^{n/p} e_n(B_{a2^{n/p}})$$

$$\lesssim \mathrm{diam}(B) + a^{1/(q-1)} \sum_{n \geq 0} 2^{-nq/(q-1)p} \sum_{0 \leq k \leq n} \left(2^{2k/p} e_k(B)\right)^{q/(q-1)}$$

$$= \mathrm{diam}(B) + a^{1/(q-1)} \sum_{k \geq 0} \sum_{n \geq k} 2^{-nq/(q-1)p} \left(2^{2k/p} e_k(B)\right)^{q/(q-1)}$$

$$\lesssim \mathrm{diam}(B) + a^{1/(q-1)} \sum_{k \geq 0} \left(2^{k/p} e_k(B)\right)^{q/(q-1)}.$$

Applying Corollary 2.7 and optimizing over $a > 0$ yields

$$\gamma_p(B) \lesssim \mathrm{diam}(B) + \left[\sum_{n \geq 0} \left(2^{n/p} e_n(B)\right)^{q/(q-1)}\right]^{(q-1)/q}.$$

It remains to note that $\mathrm{diam}(B) \leq 2e_0(B)$, so that the first term can be absorbed in the second at the expense of the universal constant. $\square$

Remark 4.4. An inspection of the proof shows that the universal constant in Theorem 4.1 blows up as $q \downarrow 1$. It would be interesting to understand whether there is an analogue of Theorem 4.1 that holds in the limiting case $q = 1$: that is, whether there is a general geometric mechanism that ensures the sharp bound

$$\gamma_p(B) \asymp \sup_{n \geq 0} 2^{n/p} e_n(B).$$

(that the right-hand side is a lower bound on $\gamma_p(B)$ is trivial). This situation is illustrated by the example of section 3.2: in this case both the assumption and the conclusion of Theorem 4.1 hold for $q = 1$ (the assumption holds by Remark 4.2 and $B_t - B_t \subseteq 2B_{\sqrt{2}t}$, while the conclusion can be deduced from [16, Exercise 2.2.15]), but Theorem 4.1 is not sufficiently sharp to capture this example.

4.2. **Uniformly convex sets.** In this section, we exhibit an important situation where the assumption of Theorem 4.1 can be verified by imposing additional geometric structure on the set B: we show that the assumption holds when B is q-convex. This recovers a fundamental result of Talagrand [16, §4.1].

Let $(X, \|\cdot\|)$ be any Banach space, and let $B \subset X$ be a symmetric convex set. As usual, we denote by $\|\cdot\|_B$ the gauge of B. We recall the following definition.

Definition 4.5. Let $q \geq 2$. A symmetric convex set B is called *q-convex* if

$$\left\| \frac{x+y}{2} \right\|_B \leq 1 - \eta \|x - y\|_B^q$$

for all $x, y \in B$, where $\eta > 0$ is an absolute constant.

We will prove the following result.

Corollary 4.6 ([16]). *Let B be a symmetric convex set in a Banach space $(X, \|\cdot\|)$, and assume that B is q-convex (with constant η). Then*

$$\gamma_p(B) \lesssim \left[\sum_{n \geq 0} \left(2^{n/p} e_n(B) \right)^{q/(q-1)} \right]^{(q-1)/q},$$

where the universal constant depends on p, q, and η only.

To connect this result to the explicit computations in section 3.1, we recall that ℓ_q-ellipsoids are $\max(2, q)$-convex [3]. This shows that the case $2 \leq q < \infty$ of Proposition 3.1 is in fact a manifestation of the much more general phenomenon described by Corollary 4.6: we emphasize that the present result requires no assumption of any kind on the norm $\|\cdot\|$. On the other hand, it is impossible for a convex set to be q-convex with $q < 2$ (Hilbert space is maximally convex), so that uniform convexity cannot explain the behavior of ℓ_q-ellipsoids for $q < 2$. We will nonetheless see in the next section that the latter case can also be understood as a manifestation of the general geometric principle described by Theorem 4.1.

We prove Corollary 4.6 by verifying the assumption of Theorem 4.1.

Lemma 4.7. *Let B be a q-convex set and $t \geq 0$. Then*

$$\|y - z\|_B^q \lesssim t \|y - z\| \quad \text{for every } y, z \in B_t,$$

where the universal constant depends on q and η only.

We will give two different proofs of this lemma. The first proof is pedestrian, but perhaps not very intuitive. The second proof is more intuitive, as it is close in spirit to the intuition developed in Remark 4.2; however, this proof requires us to use an alternative (but equivalent) formulation of the q-convexity property.

First proof. By Proposition 2.5, we have $\pi_t(y) = y$ for $y \in B_t$. Thus

$$\|y\|_B = \inf_u \{ \|u\|_B + t\|y - u\| \} \leq \left\| \frac{y+z}{2} \right\|_B + t \left\| \frac{y-z}{2} \right\|$$

for any $y, z \in B_t$. Similarly, exchanging the role of y and z, we obtain

$$1 \le \left\| \frac{y+z}{2\gamma} \right\|_B + t \left\| \frac{y-z}{2\gamma} \right\|, \qquad \gamma := \|y\|_B \vee \|z\|_B.$$

But note that $\|y/\gamma\|_B \le 1$ and $\|z/\gamma\|_B \le 1$ by the definition of γ. Therefore, applying the q-convexity assumption to the first term on the right yields

$$\|y-z\|_B^q \le \frac{\gamma^{q-1}}{2\eta} \, t \|y-z\|$$

for any $y, z \in B_t$. The proof is completed by noting that $\gamma \le 1$. $\qquad\square$

Second proof. An equivalent characterization of the q-convexity property is as follows [17, Corollary 1]: B is q-convex if and only if

$$\langle j_y - j_z, y - z \rangle \gtrsim \|y-z\|_B^q$$

for all $j_y \in J_y := \{u \in X^* : \langle u, y \rangle = \|y\|_B^q, \ \|u\|_B^* \le \|y\|_B^{q-1}\}$ and $j_z \in J_z$, where the universal constant depends on q, η only. Note that J_y is none other than the subdifferential of the map $y \mapsto \|y\|_B^q/q$ (cf. Corollary 2.8), so this characterization is rather intuitive: B is q-convex precisely when the map $y \mapsto \|y\|_B^q$ exhibits a uniform improvement over the usual first-order condition for convexity.

With this formulation in hand, the lemma follows easily. Let $y, z \in B_t$. By definition of B_t, we can choose $u_y \in X^*$ with $\langle u_y, y \rangle = \|y\|_B$, $\|u_y\|_B^* \le 1$, $\|u_y\|^* \le t$. Choose $u_z \in X^*$ analogously. Setting $j_y = u_y \|y\|_B^{q-1}$ and $j_z = u_z \|z\|_B^{q-1}$ gives

$$\|y-z\|_B^q \lesssim \langle j_y - j_z, y - z \rangle \le \|j_y - j_z\|^* \|y-z\| \le 2t\|y-z\|.$$

This completes the proof. $\qquad\square$

It is now trivial to complete the proof of Corollary 4.6.

Proof of Corollary 4.6. We may as well assume that B is compact: if B is not precompact, the right-hand side of the desired inequality is infinite and there is nothing to prove; if B is precompact, there is no loss of generality in assuming that it is also closed. It remains to apply Theorem 4.1 and Lemma 4.7. $\qquad\square$

4.3. ℓ_q-balls and unconditional bases.

We have seen in the previous section that uniform convexity cannot explain the behavior of ℓ_q-ellipsoids in Hilbert space that was observed in section 3.1. We will presently show that this behavior is nonetheless a manifestation of the general geometric principle of Theorem 4.1. It will follow immediately that the same behavior persists in a much larger family of Banach spaces (but not in a setting as general as for q-convex sets).

To understand what is going on, let us take inspiration from the second proof of Lemma 4.7 (and from Remark 4.2). For any $x \in X$, choose any point $j_x \in X^*$ be such that $\langle j_x, x \rangle = \|x\|_B^q$ and $\|j_x\|_B^* \le \|x\|_B^{q-1}$. As

$$\|y-z\|_B^q = \langle j_{y-z}, y - z \rangle \le \|j_{y-z}\|^* \|y-z\|,$$

the assumption of Theorem 4.1 would follow if we could show that $\|j_{y-z}\|^* \lesssim t$ whenever $y, z \in B_t$. We can always choose $\|j_x\|^* \le t$ when $x \in B_t$, but this does not in itself yield the desired result: $y, z \in B_t$ does not imply $y - z \in B_t$.

To obtain the desired bound, we must find a relation between j_{y-z} and j_y, j_z. The q-convexity assumption provides the inequality $\langle j_{y-z}, y - z \rangle \lesssim \langle j_y - j_z, y - z \rangle$, which is particularly convenient for this purpose. However, this is by no means the only way to achieve our goal. In the case of ℓ_q-ellipsoids, we will use a completely different

geometric property: in this case we observe that $|j_{y-z}| \lesssim |j_y| + |j_z|$ coordinatewise. This simple device allows us to reach the same conclusion as in the q-convex case as long as the dual norm $\|\cdot\|^*$ respects the coordinatewise ordering.

We proceed to make this idea precise. We first recall the class of Banach spaces that possess the desired monotonicity properties [1, §3.1].

Definition 4.8. Let $(X, \|\cdot\|)$ be a Banach space and let $\{e_n\}$ be a basis for X. The basis is said to be unconditional with constant K if

$$\left\| \sum_{n=1}^{N} a_n e_n \right\| \leq K \left\| \sum_{n=1}^{N} b_n e_n \right\|$$

for all $N \in \mathbb{N}$ and scalars $a_n, b_n \in \mathbb{R}$ such that $|a_n| \leq |b_n|$ for all n.

We recall for future reference that if $\{e_n\}$ is an unconditional basis in X with constant K, then the biorthogonal sequence $\{e_n^*\}$ is an unconditional basic sequence in X^* with the same constant K [1, Proposition 3.2.1].

Remark 4.9. The notion of a K-unconditional basis is often defined in a slightly different way than we have done above: a basis is unconditional with constant K if

$$\left\| \sum_{n=1}^{N} \varepsilon_n b_n e_n \right\| \leq K \left\| \sum_{n=1}^{N} b_n e_n \right\|$$

for all $N \in \mathbb{N}$, $b_n \in \mathbb{R}$, and $\varepsilon_n \in \{-1, +1\}$, that is, if the norm of $\sum_{n=1}^{N} b_n e_n$ is approximately invariant to sign changes of the coefficients b_n. The more general property of Definition 4.8 is however readily deduced from this alternative definition (for example, by choosing random signs ε_n such that $a_n = \mathbf{E}[\varepsilon_n b_n]$).

In the following let $(X, \|\cdot\|)$ be a Banach space and let $\{e_n\}$ be an unconditional basis with constant K. Fix $1 < q < \infty$, and define the ℓ_q-ball $B \subset X$ as follows:

$$B = \left\{ \sum_{i=1}^{d} z_i e_i : \sum_{i=1}^{d} |z_i|^q \leq 1 \right\}$$

(our result will be independent of d, and therefore extends readily to infinite dimension). Note that the ℓ_q-ellipsoids considered in section 3.1 correspond to the special case where $\{e_i\}$ is the standard basis in $\mathbb{R}^d$ and $\|x\|^2 = \sum_i b_i^2 x_i^2$.

Corollary 4.10. *In the setting of this section, we have*

$$\gamma_p(B) \lesssim \left[\sum_{n \geq 0} \left(2^{n/p} e_n(B) \right)^{q/(q-1)} \right]^{(q-1)/q},$$

where the universal constant depends on p, q, and K only.

Proof. The norm $\|\cdot\|$ on X can be transferred to $\mathbb{R}^d$ by defining $\|z\| := \|\sum_{i=1}^{d} z_i e_i\|$ for $z \in \mathbb{R}^d$. There is therefore no loss of generality in assuming that $X = \mathbb{R}^d$ with the above norm, that $\{e_i\} = \{e_i^*\}$ is the standard basis, and that $\|x\|_B$ is the ℓ_q-norm on $\mathbb{R}^d$, as we will do in the sequel for notational simplicity. (We emphasize, however, that $\|\cdot\|$ is *not* the Euclidean norm, so that the present setting does not reduce to the Euclidean setting considered previously in section 3.1).

As $\|x\|_B$ is the ℓ_q-norm, we can compute

$$\frac{\partial\|x\|_B}{\partial x_i} = \frac{|x_i|^{q-1}}{\|x\|_B^{q-1}}\operatorname{sign}(x_i).$$

By Corollary 2.8, we can write

$$B_t = \{x \in B : \||x|^{q-1}\operatorname{sign}(x)\|^* \le t\|x\|_B^{q-1}\}.$$

Now note that for any vectors $x, y \in \mathbb{R}^d$, we have

$$\|x - y\|_B^q = \langle |x-y|^{q-1}\operatorname{sign}(x-y), x-y\rangle \le \||x-y|^{q-1}\operatorname{sign}(x-y)\|^*\|x-y\|.$$

Moreover, as $|x - y|^{q-1} \le 2^{(q-2)_+}(|x|^{q-1} + |y|^{q-1})$, we have

$$\||x-y|^{q-1}\operatorname{sign}(x-y)\|^* \le 2^{(q-2)_+}K\||x|^{q-1} + |y|^{q-1}\|^* \le 2^{1+(q-2)_+}K^2 t$$

for all $x, y \in B_t$ using the unconditional property of the dual basis $\{e_n^*\}$. Thus

$$\|x - y\|_B^q \le 2^{1+(q-2)_+}K^2 t\|x - y\|$$

whenever $x, y \in B_t$, and it remains to invoke Theorem 4.1. $\qquad\square$

We have now given two distinct explanations for the behavior of ℓ_q-ellipsoids observed in section 3.1. When $q \ge 2$, such sets are q-convex and the result follows from the general principle described by Corollary 4.6. In this setting, the result remains valid when $\|\cdot\|$ is an arbitrary norm. When $q < 2$, the observed behavior is described by Corollary 4.10, which exploits a more special geometric property of ℓ_q-balls. In this setting, the result also remains valid for a large class of norms $\|\cdot\|$, but we require the additional restriction that the underlying basis is unconditional. It appears that these two cases possess a genuinely different geometry, which is completely hidden in the statement of Proposition 3.1.

Acknowledgments. The author would like to thank the anonymous referees for helpful comments that improved the presentation of this paper.

References

[1] F. Albiac and N. J. Kalton. *Topics in Banach space theory*, volume 233 of *Graduate Texts in Mathematics*. Springer, New York, 2006.

[2] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[3] B. Beauzamy. *Introduction to Banach spaces and their geometry*, volume 68 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam-New York, 1982.

[4] C. Bennett and R. Sharpley. *Interpolation of operators*, volume 129 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1988.

[5] B. Carl. Entropy numbers of diagonal operators with an application to eigenvalue problems. *J. Approx. Theory*, 32(2):135–150, 1981.

[6] R. A. DeVore and V. A. Popov. Interpolation spaces and nonlinear approximation. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 191–205. Springer, Berlin, 1988.

[7] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1:290–330, 1967.

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

[9] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1993.

[10] A. N. Kolmogorov and V. M. Tihomirov. ε-entropy and ε-capacity of sets in function spaces. *Uspehi Mat. Nauk*, 14(2 (86)):3–86, 1959.

[11] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin, 1991.

[12] J. Peetre and G. Sparr. Interpolation of normed abelian groups. *Ann. Mat. Pura Appl. (4)*, 92:217–262, 1972.

[13] A. Pietsch. Approximation spaces. *J. Approx. Theory*, 32(2):115–134, 1981.

[14] V. N. Sudakov. Gauss and Cauchy measures and ε-entropy. *Dokl. Akad. Nauk SSSR*, 185:51–53, 1969.

[15] M. Talagrand. Majorizing measures: the generic chaining. *Ann. Probab.*, 24(3):1049–1103, 1996.

[16] M. Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, Heidelberg, 2014.

[17] H. K. Xu. Inequalities in Banach spaces with applications. *Nonlinear Anal.*, 16(12):1127–1138, 1991.

Sherrerd Hall Room 227, Princeton University, Princeton, NJ 08544, USA
E-mail address: rvan@princeton.edu