The Annals of Probability
2018, Vol. 46, No. 3, 1764–1805
https://doi.org/10.1214/17-AOP1214

CHAINING, INTERPOLATION AND CONVEXITY II: THE CONTRACTION PRINCIPLE[1]

BY RAMON VAN HANDEL

Princeton University

The generic chaining method provides a sharp description of the suprema of many random processes in terms of the geometry of their index sets. The chaining functionals that arise in this theory are however notoriously difficult to control in any given situation. In the first paper in this series, we introduced a particularly simple method for producing the requisite multiscale geometry by means of real interpolation. This method is easy to use, but does not always yield sharp bounds on chaining functionals. In the present paper, we show that a refinement of the interpolation method provides a canonical mechanism for controlling chaining functionals. The key innovation is a simple but powerful contraction principle that makes it possible to efficiently exploit interpolation. We illustrate the utility of this approach by developing new dimension-free bounds on the norms of random matrices and on chaining functionals in Banach lattices. As another application, we give a remarkably short interpolation proof of the majorizing measure theorem that entirely avoids the greedy construction that lies at the heart of earlier proofs.

1. Introduction. The development of sharp bounds on the suprema of random processes is of fundamental importance in diverse areas of pure and applied mathematics. Such problems arise routinely, for example, in probability theory, functional analysis, convex geometry, mathematical statistics and theoretical computer science.

It has long been understood that the behavior of suprema of random processes is intimately connected with the geometry of their index sets. This idea has culminated in a remarkably general theory due to M. Talagrand that captures the precise connection between the underlying probabilistic and geometric structures for many interesting types of random processes. For example, the classic result in this theory, known (for historical reasons) as the majorizing measure theorem, provides a sharp geometric characterization of the suprema of Gaussian processes.

Received November 2016; revised July 2017.

[1]Supported in part by NSF Grant CAREER-DMS-1148711 and by the ARO through PECASE award W911NF-14-1-0094.

MSC2010 subject classifications. 60B11, 60G15, 41A46, 46B20, 46B70.

Key words and phrases. Generic chaining, majorizing measures, entropy numbers, real interpolation, suprema of random processes, random matrices.

THEOREM 1.1 ([11]). *Let $(X_x)_{x\in T}$ be a centered Gaussian process and let $d(x,y) = (\mathbf{E}|X_x - X_y|^2)^{1/2}$ be the associated natural metric on T. Then*

$$\mathbf{E}\Big[\sup_{x\in T} X_x\Big] \asymp \gamma_2^*(T) := \inf \sup_{x\in T} \sum_{n\geq 0} 2^{n/2} d(x, T_n),$$

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

The method behind the proof of Theorem 1.1 is called the generic chaining. It is by no means restricted to the setting of Gaussian processes, and analogues of Theorem 1.1 exist in various settings. We refer to the monograph [11] for a comprehensive treatment of this theory and its applications.

The majorizing measure theorem provides in principle a complete geometric understanding (up to universal constants) of the suprema of Gaussian processes. The chaining functional $\gamma_2^*(T)$ captures the relevant geometric structure: it quantifies how well the index set T can be approximated, in a multiscale fashion, by increasingly fine discrete nets T_n. The apparently definitive nature of Theorem 1.1 belies the fact that this result is often very difficult to use in any concrete situation. The problem is that while Theorem 1.1 guarantees that there must exist some optimal sequence of nets T_n that yields a sharp bound on the supremum of any given Gaussian process, the theorem does not explain how to find such nets. In many cases, straightforward discretization of the index set (Dudley's inequality) gives rise to suboptimal bounds, and it is not clear how such bounds can be improved.

Even without going beyond the setting of Gaussian processes, there are plenty of challenging problems, for example, in random matrix theory [9, 13, 14], that remain unsolved due to the lack of understanding of how to control the supremum of some concrete Gaussian process; in fact, even in cases where the supremum of a Gaussian process can be trivially bounded by probabilistic means, the underlying geometry often remains a mystery; cf. [11], page 50. From this perspective, the generic chaining theory remains very far from being well understood. It is therefore of considerable interest to develop new mechanisms for the control of chaining functionals such as $\gamma_2^*(T)$. The aim of this paper is to introduce some new ideas in this direction.

The main (nontrivial) technique that has been used to date to control chaining functionals is contained in the proof of Theorem 1.1. To show that $\gamma_2^*(T)$ is bounded above by the expected supremum of a Gaussian process, a sequence of nets T_n is constructed by repeatedly partitioning the set T in a greedy fashion, using the functional $G(A) := \mathbf{E}[\sup_{x\in A} X_x]$ to quantify the size of each partition element. It is necessary to carefully select the partition elements at each stage of the construction in order to control future iterations, which requires fairly delicate arguments (cf. [11], Section 2.6). It turns out, however, that the proof does not rely heavily on special properties of Gaussian processes: the only property of the functional $G(A)$ that is used is that a certain "growth condition" is satisfied. If one can design another functional $F(A)$ that mimics this property of Gaussian processes, then the same proof yields an upper bound on $\gamma_2^*(T)$ in terms of $F(T)$.

In principle, this partitioning scheme provides a canonical method for bounding chaining functionals such as $\gamma_2^*(T)$: it is always possible to choose a functional satisfying the requisite growth condition that gives a sharp bound on $\gamma_2^*(T)$. This observation has little practical relevance, as this conclusion follows from the fact that the chaining functional itself satisfies the growth condition (cf. [11], pages 38–40) which does not help to obtain explicit bounds on these functionals. Nonetheless, this observation shows that no loss is incurred in the partitioning scheme per se, so that its application is only limited by our ability to design good growth functionals that admit explicit bounds. Unfortunately, the latter requires considerable ingenuity, and has been carried out successfully in a limited number of cases.

In the first paper in this series [12], the author introduced a new method to bound chaining functionals that is inspired by real interpolation of Banach spaces. The technique developed in [12] is completely elementary and is readily amenable to explicit computations, unlike the growth functional method. This approach considerably simplifies and clarifies some of the most basic ideas in the generic chaining theory, such as the construction of chaining functionals on uniformly convex bodies. On the other hand, this basic method is not always guaranteed to give sharp bounds on $\gamma_2^*(T)$, as can be seen in simple examples (cf. [12], Section 3.3). It is therefore natural to expect that the utility of the interpolation method may be restricted to certain special situations whose geometry is well captured by this construction.

The key insight of the present paper is that this is not the case: interpolation provides a canonical method for bounding chaining functionals. The problem with the basic method of [12] does not lie with the interpolation method itself, but is rather due to the fact that this method was inefficiently exploited in its simplest form. What is missing is a simple but apparently fundamental ingredient, a contraction principle, that will be developed and systematically exploited in this paper. Roughly speaking, the contraction principle states that one can control chaining functionals such as $\gamma_2^*(T)$ whenever one has suitable control on the entropy numbers of all subsets $A \subseteq T$. A precise statement of this principle will be given in Section 3 below, and its utility will be illustrated throughout the rest of paper.

The combination of the interpolation method and the contraction principle provides a foundation for the generic chaining theory that yields significantly simpler proofs and appears to be easier to use (at least in this author's opinion) than the classical approach through growth functionals. This approach will be illustrated in a number of old and new applications. For example, we will fully recover the majorizing measure theorem with a remarkably short proof that does not involve any greedy partitioning scheme. The latter is surprising in its own right, as a greedy construction lies very much at the core of earlier proofs of Theorem 1.1.

This paper is organized as follows. In Section 2, we set up the basic definitions and notation that will be used throughout the paper. Section 3 develops the main idea of this paper, the contraction principle. This principle is first illustrated by means of some elementary examples in Section 4, which includes a brief review of the basic interpolation method developed in [12]. In Section 5, we develop a geometric principle that resolves a question posed in [12], Remark 4.4. We then use this principle to develop new results on the behavior of chaining functionals on Banach lattices, as well as to recover classical results on uniformly convex bodies. In Section 6, we develop a very simple proof of Theorem 1.1 using the machinery of this paper. We also show that the growth functional machinery that lies at the heart of [11] can be recovered as a special case of our approach. Finally, in Section 7 we exploit the methods of this paper to develop new dimension-free bounds on the operator norms of structured random matrices.

2. Basic definitions and notation. The aim of this section is to set up the basic definitions and notation that will be used throughout the paper. We introduce a general setting that will be specialized to different problems as needed in the sequel.

Let (X, d) be a metric space. We begin by defining entropy numbers.

DEFINITION 2.1. For every $A \subseteq X$ and $n \geq 0$, define the *entropy number*

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

(In this definition, the net $S \subseteq X$ is not required to be a subset of A.)

Another way to interpret $e_n(A)$ is by noting that A can be covered by less than 2^{2^n} balls of radius $e_n(A)$. It is useful to recall a classical observation (the duality between covering and packing) that will be needed below.

LEMMA 2.2. *Let $n \geq 0$ and $N = 2^{2^n}$. Then for every $0 < \delta < e_n(A)$, there exist points $x_1, \dots, x_N \in A$ such that $d(x_i, x_j) > \delta$ for all $i \neq j$.*

PROOF. Select points $x_1, x_2, \dots$ as follows: choose $x_1 \in A$ arbitrarily, and choose $x_i \in A$ so that $d(x_i, x_j) > \delta$ for all $j < i$. Suppose this construction terminates in round M, that is, there does not exist $x \in A$ so that $d(x, x_j) > \delta$ for all $j \leq M$. Then setting $S = \{x_1, \dots, x_M\}$, we have $\sup_{x \in A} d(x, S) \leq \delta$. Thus $M \geq N$, as otherwise $e_n(A) \leq \delta$ which contradicts our assumption. □

We now turn to the definition of chaining functionals. For the purposes of the present paper, it will be convenient to use a slightly different definition than in Theorem 1.1 that uses partitions rather than nets. We also formulate a more general class of chaining functionals that are useful in the general generic chaining theory (beyond the setting of Gaussian processes); cf. [11].

DEFINITION 2.3. Let $T \subseteq X$. An *admissible sequence* of T is an increasing sequence $(\mathcal{A}_n)$ of partitions of T such that $|\mathcal{A}_n| < 2^{2^n}$ for all $n \geq 0$. For every $x \in T$, we denote by $A_n(x)$ the unique element of $\mathcal{A}_n$ that contains x.

DEFINITION 2.4. Let $T \subseteq X$. For $\alpha > 0$ and $p \geq 1$, define

$$\gamma_{\alpha,p}(T) := \left[\inf \sup_{x \in T} \sum_{n \geq 0} \big(2^{n/\alpha} \operatorname{diam}(A_n(x))\big)^p\right]^{1/p},$$

where the infimum is taken over all admissible sequences of T. The most important case $p = 1$ is denoted as $\gamma_\alpha(T) := \gamma_{\alpha,1}(T)$.

It is an easy fact that the chaining functional $\gamma_2^*(T)$ that appears in Theorem 1.1 satisfies $\gamma_2^*(T) \leq \gamma_2(T)$: given an admissible sequence $(\mathcal{A}_n)$ of T, we may simply select a net T_n by choosing one point arbitrarily in every element of the partition $\mathcal{A}_n$. As our interest in this paper is to obtain upper bounds on $\gamma_2(T)$, these trivially give upper bounds on $\gamma_2^*(T)$ as well. It is not difficult to show that these quantities are actually always of the same order; cf. [11], Section 2.3. This will also follow as a trivial application of the main result of this paper; see Section 4.1 below.

Let us emphasize that the definitions of $e_n(A)$ and $\gamma_{\alpha,p}(T)$ depend on the metric of the underlying metric space (X, d). In some situations, we will be working with multiple metrics; in this case, the metric d that is used to define the above quantities will be denoted explicitly by writing $e_n(A, d)$, $\gamma_{\alpha,p}(T, d)$ and $\operatorname{diam}(A, d)$.

We will write $a \lesssim b$ if $a \leq Cb$ for a universal constant C, and we write $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. In cases where the universal constant depends on some parameter of the problem, this will be indicated explicitly.

3. The contraction principle. At the heart of this paper lies a simple but apparently fundamental principle that will be developed in this section. The basic idea is that we can control the chaining functionals $\gamma_{\alpha,p}(T)$ when we have suitable control on the entropy numbers $e_n(A)$ of all subsets $A \subseteq T$.

THEOREM 3.1 (Contraction principle). *Let $s_n(x) \geq 0$ and $a \geq 0$ satisfy*

$$e_n(A) \leq a \operatorname{diam}(A) + \sup_{x \in A} s_n(x)$$

for every $n \geq 0$ and $A \subseteq T$. Then

$$\gamma_{\alpha,p}(T) \lesssim a\gamma_{\alpha,p}(T) + \left[\sup_{x \in T} \sum_{n \geq 0} \big(2^{n/\alpha} s_n(x)\big)^p\right]^{1/p},$$

where the universal constant depends only on α.

Of course, this result is of interest only when a can be chosen sufficiently small, in which case it immediately yields an upper bound on $\gamma_{\alpha,p}(T)$.

It should be emphasized that the only nontrivial aspect of Theorem 3.1 is to discover the correct formulation of this principle; no difficulties are encountered in the proof. What may be far from obvious at present is that this is in fact a powerful or even useful principle. This will become increasingly clear in the following sections, where we will see that the interpolation method of [12] provides a canonical mechanism for generating controls $s_n(x)$.

As Theorem 3.1 lies at the core of this paper, we give two slightly different proofs. The first proof explains the term "contraction principle." The second proof is instructive for contrasting our approach with that of [11].

3.1. *First proof.* The idea of the proof is that the assumption of Theorem 3.1 allows us to construct from any admissible sequence a new admissible sequence that provides more control on the value of the chaining functional.

FIRST PROOF OF THEOREM 3.1. As $e_n(A) \le \operatorname{diam}(T)$ for every $A \subseteq T$, we can assume without loss of generality that $s_n(x) \le \operatorname{diam}(T)$ for all n, x.

Let $(\mathcal{A}_n)$ be an admissible sequence of T. For every $n \ge 1$ and partition element $A_n \in \mathcal{A}_n$, we construct sets A_n^{ij} as follows. We first partition A_n into n segments (here $1 \le i < n$):

$$A_n^i := \{x \in A_n : 2^{-2i/\alpha} \operatorname{diam}(T) < s_n(x) \le 2^{-2(i-1)/\alpha} \operatorname{diam}(T)\},$$
$$A_n^n := \{x \in A_n : s_n(x) \le 2^{-2(n-1)/\alpha} \operatorname{diam}(T)\}.$$

The point of this step is to ensure that for all $x \in A_n^i$ and $i \le n$, we have

$$\sup_{y \in A_n^i} s_n(y) \le 2^{2/\alpha} s_n(x) + 2^{-2(n-1)/\alpha} \operatorname{diam}(T),$$

that is, $s_n(x)$ is nearly constant on A_n^i. Using the assumption of the theorem, we can further partition each A_n^i into less than 2^{2^n} pieces A_n^{ij} such that

$$\operatorname{diam}(A_n^{ij}) \le 2a \operatorname{diam}(A_n) + 2^{1+2/\alpha} s_n(x) + 2^{1-2(n-1)/\alpha} \operatorname{diam}(T)$$

for all $x \in A_n^{ij}$. Let $\mathcal{C}_{n+3}$ be the partition generated by all sets A_k^{ij}, $k \le n$, i, j thus constructed. Then $|\mathcal{C}_{n+3}| < \prod_{k=1}^n k(2^{2^k})^2 < 2^{2^{n+3}}$. Defining $\mathcal{C}_k = \{T\}$ for $0 \le k \le 3$, we readily obtain

$$\begin{aligned}\gamma_{\alpha,p}(T) &\le \left[\sup_{x\in T} \sum_{n\ge 0} (2^{n/\alpha} \operatorname{diam}(C_n(x)))^p\right]^{1/p} \\ &\lesssim a\left[\sup_{x\in T} \sum_{n\ge 0} (2^{n/\alpha} \operatorname{diam}(A_n(x)))^p\right]^{1/p} \\ &\quad + \left[\sup_{x\in T} \sum_{n\ge 0} (2^{n/\alpha} s_n(x))^p\right]^{1/p} + \operatorname{diam}(T),\end{aligned}$$

where the universal constant depends only on α. The last term on the right can be absorbed in the first two as $\mathrm{diam}(T) \le 2e_0(T) \le 2a\,\mathrm{diam}(A_0(x)) + 2\sup_{x\in T} s_0(x)$. It remains to note that the admissible sequence $(\mathcal{A}_n)$ was arbitrary, so we can take the infimum over $(\mathcal{A}_n)$ on the right-hand side. $\square$

One way to interpret this proof is as follows. We used the assumption of Theorem 3.1 to define a mapping $\Gamma : \mathcal{A} \mapsto \mathcal{C}$ that assigns to every admissible sequence $\mathcal{A} = (\mathcal{A}_n)$ a new admissible sequence $\mathcal{C} = (\mathcal{C}_n)$. This mapping can be thought of as inducing a form of dynamics on the space of admissible sequences. If we define the value of $\mathcal{A}$ and the target upper bound as

$$\mathrm{val}(\mathcal{A}) = \left[\sup_{x\in T} \sum_{n\ge 0} \big(2^{n/\alpha}\,\mathrm{diam}\big(A_n(x)\big)\big)^p\right]^{1/p},$$

$$S = \left[\sup_{x\in T} \sum_{n\ge 0} \big(2^{n/\alpha} s_n(x)\big)^p\right]^{1/p},$$

then we have shown in the proof that

$$\mathrm{val}\big(\Gamma(\mathcal{A})\big) \le Ca\,\mathrm{val}(\mathcal{A}) + CS$$

for a universal constant C. If a can be chosen so that $Ca < 1$, then Γ defines a sort of contraction on the space of admissible sequences. This ensures that there exists an admissible sequence with value $\mathrm{val}(\mathcal{A}) \lesssim S$, which is the conclusion we seek. This procedure is reminiscent of the contraction mapping principle, which is why we call Theorem 3.1 the contraction principle.

3.2. *Second proof.* The above proof ensures the existence of a good admissible partition without directly constructing this partition. This is in contrast to the partitioning scheme of [11], where a good admissible partition is explicitly constructed in the proof. We presently show that by organizing the proof in a slightly different way, we also obtain an explicit construction.

SECOND PROOF OF THEOREM 3.1. We construct an increasing sequence of partitions $(\mathcal{B}_n)$ of T by induction. First, set $\mathcal{B}_0 = \{T\}$. Now suppose partitions $\mathcal{B}_0, \dots, \mathcal{B}_{n-1}$ have already been constructed. We first split every set $B_{n-1} \in \mathcal{B}_{n-1}$ into n segments (here $1 \le i < n$):

$$B_n^i := \big\{x \in B_{n-1} : 2^{-2i/\alpha}\,\mathrm{diam}(T) < s_n(x) \le 2^{-2(i-1)/\alpha}\,\mathrm{diam}(T)\big\},$$

$$B_n^n := \big\{x \in B_{n-1} : s_n(x) \le 2^{-2(n-1)/\alpha}\,\mathrm{diam}(T)\big\},$$

and then further subdivide each B_n^i into less than 2^{2^n} pieces B_n^{ij} so that

$$\mathrm{diam}\big(B_n^{ij}\big) \le 2a\,\mathrm{diam}(B_{n-1}) + 2^{1+2/\alpha} s_n(x) + 2^{1-2(n-1)/\alpha}\,\mathrm{diam}(T)$$

for all $x \in B_n^{ij}$. Now let $\mathcal{B}_n = \{B_n^{ij} : B_{n-1} \in \mathcal{B}_{n-1}, i \le n, j < 2^{2^n}\}$. As $|\mathcal{B}_n| < n2^{2^n}|\mathcal{B}_{n-1}| < \prod_{k=1}^n k2^{2^k} < 2^{2^{n+2}}$, $(\mathcal{B}_n)$ is not itself an admissible sequence. We can however easily convert it to an admissible sequence $(\mathcal{A}_n)$ by defining $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$ and $\mathcal{A}_{n+2} = \mathcal{B}_n$.

Now note that by construction, we have

$$\operatorname{diam}(A_n(x)) \le 2a \operatorname{diam}(A_{n-1}(x)) + 2^{1+2/\alpha} s_{n-2}(x) + 2^{1-2(n-3)/\alpha} \operatorname{diam}(T)$$

whenever $n \ge 3$. Therefore, in the notation of the previous subsection,

$$\operatorname{val}(\mathcal{A}) \le Ca \operatorname{val}(\mathcal{A}) + CS$$

for a universal constant C depending only on α, where we used the same argument as in the first proof of Theorem 3.1 to absorb the $\operatorname{diam}(T)$ term. If $Ca \le \frac{1}{2}$, say, then we obtain the desired bound $\gamma_{\alpha,p}(T) \le \operatorname{val}(\mathcal{A}) \le 2CS$ (this is the interesting case). On the other hand, if $Ca > \frac{1}{2}$, we trivially have $\gamma_{\alpha,p}(T) \le 2Ca\gamma_{\alpha,p}(T)$. Thus the conclusion of Theorem 3.1 follows. □

The second proof of Theorem 3.1 is reminiscent of the partitioning scheme of [11] to the extent that an admissible sequence is constructed by repeatedly partitioning the index set T. In contrast to the method of [11], however, the present approach is completely devoid of subtlety: the partitioning at each stage is performed in the most naive possible way by breaking up each set arbitrarily into pieces of the smallest possible diameter. We will nonetheless see in Section 6 that the growth functional machinery of [11] can be recovered from Theorem 3.1 with a remarkably simple proof. In our approach, the growth functional will play no role in the partitioning process itself, but will only be used to produce controls $s_n(x)$ that yield good *a priori* bounds on the entropy numbers $e_n(A)$ for $A \subseteq T$.

REMARK 3.2. The chaining functionals $\gamma_{\alpha,p}(T, d)$ that we consider in this paper use a single distance d to control the diameter of the admissible sequence at every scale. Applications of the generic chaining theory to canonical or infinitely divisible processes require a more general form of the chaining functionals that allows one to use a different distance d_n for each scale n ([11], Chapter 10). The contraction principle, as well as the interpolation machinery developed in the following sections (particularly Section 6.2), extend readily to this setting. However, as such functionals will not be used in any of the applications of this paper, we have chosen to formulate our results in the more basic setting for notational and conceptual simplicity.

4. Simple illustrations. Before we can apply Theorem 3.1 in a nontrivial manner, we should develop some insight into the meaning of the numbers $s_n(x)$ and basic ways in which they can be constructed. To this end, we aim in this section to illustrate Theorem 3.1 in the simplest cases. All results developed here admit more direct proofs, but the present treatment is intended to help understand the meaning of Theorem 3.1.

4.1. *Admissible sequences and nets.* When the abstract statement of Theorem 3.1 is first encountered, it may be far from obvious why

$$e_n(A) \le a\,\mathrm{diam}(A) + \sup_{x\in A} s_n(x)$$

is a natural assumption. The relevance of the numbers $s_n(x)$ can be immediately clarified by observing that a canonical choice is already built into the definition of the chaining functional $\gamma_{\alpha,p}(T)$.

LEMMA 4.1. *Let $(\mathcal{A}_n)$ be any admissible sequence of T. Then the choice $s_n(x) = \mathrm{diam}(A_n(x))$ satisfies the assumption of Theorem 3.1 with $a = 0$.*

PROOF. Any $A \subseteq T$ is covered by less than 2^{2^n} sets $\{A_n(x) : x \in A\}$ of diameter at most $\sup_{x\in A} \mathrm{diam}(A_n(x))$, so $e_n(A) \le \sup_{x\in A} \mathrm{diam}(A_n(x))$. □

Of course, with this choice, Theorem 3.1 yields $\gamma_{\alpha,p}(T) \lesssim \gamma_{\alpha,p}(T)$ which is not very interesting. Nonetheless, Lemma 4.1 explains why bounding entropy numbers $e_n(A)$ in terms of controls $s_n(x)$ is entirely natural. Moreover, we see that Theorem 3.1 can in principle always give a sharp bound on $\gamma_{\alpha,p}(T)$.

As an only slightly less trivial example, let us show that the functional $\gamma_2^*(T)$ defined in the Introduction is always of the same order as $\gamma_2(T)$.

LEMMA 4.2. $\gamma_2^*(T) \asymp \gamma_2(T)$.

PROOF. As was noted in Section 2, the inequality $\gamma_2^*(T) \le \gamma_2(T)$ is trivial. To prove the converse inequality, let T_n be arbitrary sets of cardinality $|T_n| < 2^{2^n}$, and define $s_n(x) = d(x, T_n)$. The definition of entropy numbers instantly yields $e_n(A) \le \sup_{x\in A} s_n(x)$. Applying Theorem 3.1 with $a = 0$ yields

$$\gamma_2(T) \lesssim \sup_{x\in T} \sum_{n\ge 0} 2^{n/2} d(x, T_n).$$

Taking the infimum over all choices of T_n yields $\gamma_2(T) \le \gamma_2^*(T)$. □

So far, we only used Theorem 3.1 with $a = 0$ and did not exploit the "contraction" part of the contraction principle. We now provide a first illustration of an improvement that can be achieved by exploiting contraction.

4.2. *A local form of Dudley's inequality.* The most naive bound on $\gamma_2^*(T)$ is obtained by moving the supremum in its definition inside the sum. This yields the following result, which is known as Dudley's inequality:

$$\gamma_2^*(T) \le \sum_{n\ge 0} 2^{n/2} e_n(T).$$

Dudley's inequality represents the simplest possible construction where each net T_n in the definition of $\gamma_2^*(T)$ is distributed as uniformly as possible over the index set T. Unfortunately, such a simple construction proves to be suboptimal already in some of the simplest examples (cf. [11, 12]). To attain the sharp bound that is guaranteed by Theorem 1.1, it is essential to allow for the nets T_n to be constructed in a genuinely multiscale fashion. Nonetheless, Dudley's inequality is widely used in practice due to the ease with which it lends itself to explicit computations. It is no surprise that Dudley's inequality is trivially recovered by Theorem 3.1.

LEMMA 4.3. *For a universal constant C depending only on α, we have*

$$\gamma_{\alpha,p}(T) \le C\left[\sum_{n\ge 0}\big(2^{n/\alpha}e_n(T)\big)^p\right]^{1/p}.$$

As $e_n(A) \le e_n(T)$, this follows using $s_n(x) = e_n(T)$ and $a = 0$ in Theorem 3.1. However, without much additional effort, we can do slightly better using a simple application of the "contraction" part of the contraction principle.

To exploit contraction, we note that if $e_n(A) \le a\,\mathrm{diam}(A) = r$, then the assumption of Theorem 3.1 is automatically satisfied; thus the numbers $s_n(x)$ only need to control the situation where this condition fails. As A is contained in a ball of radius $\mathrm{diam}(A)$, this condition essentially means that a certain ball of radius r/a can be covered by less than 2^{2^n} balls of proportional radius r, which is a sort of doubling condition on the metric space (T, d). Let us consider the largest radius of a ball that is centered at a given point x for which this doubling condition fails:

$$e_n^{a,x}(T) := \sup\{r : e_n\big(T \cap B(x, r/a)\big) > r\},$$

where $B(x,r) := \{y \in X : d(x, y) \le r\}$. Clearly, $e_n^{a,x}(T) \le e_n(T)$, so this quantity can be viewed as a local improvement on the notion of entropy numbers. We can now use Theorem 3.1 to show that Dudley's inequality remains valid if we replace the (global) entropy numbers by their local counterparts.

LEMMA 4.4. *For universal constants C, a depending only on α, we have*

$$\gamma_{\alpha,p}(T) \le C\left[\sup_{x\in T}\sum_{n\ge 0}\big(2^{n/\alpha}e_n^{a,x}(T)\big)^p\right]^{1/p}.$$

PROOF. Let $a > 0$, $n \ge 0$ and $x \in A \subseteq T$. If $\mathrm{diam}(A) > e_n^{a,x}(T)/a$, then

$$e_n(A) \le e_n\big(T \cap B\big(x, \mathrm{diam}(A)\big)\big) \le a\,\mathrm{diam}(A)$$

by definition. On the other hand, if $\mathrm{diam}(A) \le e_n^{a,x}(T)/a$, then trivially

$$e_n(A) \le \mathrm{diam}(A) \le \frac{e_n^{a,x}(T)}{a}.$$

Thus the assumption of Theorem 3.1 holds with $s_n(x) = e_n^{a,x}(T)/a$. The proof is readily concluded by choosing a to be a small universal constant. □

An almost identical proof yields a variant of Lemma 4.4 given in [10], equation (1.9), that uses a regularized form of the local entropy numbers $e_n^{a,x}(T)$.[2] While these bounds can improve on Dudley's inequality in some esoteric (ultrametric) examples, they are not particularly useful in practice. The reason that Lemma 4.4 is included here is to help provide some initial intuition for how one might use the "contraction" part of the contraction principle. The real power of the contraction principle will however arise when it is combined with the interpolation method of [12].

4.3. *The simplest interpolation estimate.* As the interpolation method will play a crucial role in the remainder of this paper, we must begin by recalling the main idea behind this method. The aim of this section is to provide a first illustration of how interpolation can be used to generate the controls $s_n(x)$ in Theorem 3.1 by recovering the main result of [12].

The interpolation method is based on the following construction. Given a penalty function $f : X \to \mathbb{R}_+ \cup \{+\infty\}$, define the interpolation functional

$$K(t,x) := \inf_{y\in X}\bigl\{f(y) + td(x,y)\bigr\}.$$

We will assume for simplicity that the infimum in this definition is attained for each $t \ge 0$ and $x \in T$, and denote by $\pi_t(x)$ an arbitrary choice of minimizer (if the infimum is not attained we can easily modify our results to work with near-minimizers). We now define the interpolation sets

$$K_t := \bigl\{\pi_t(x) : x \in T\bigr\}.$$

The idea of the interpolation method is that the sets K_t provide a multiscale approximation of T precisely of the form suggested by Theorem 1.1.

LEMMA 4.5. *For every $a > 0$, we have*

$$\sup_{x\in T}\sum_{n\ge 0} 2^{n/\alpha} d\bigl(x, \pi_{a2^{n/\alpha}}(x)\bigr) \lesssim \frac{1}{a}\sup_{x\in T} f(x),$$

where the universal constant depends only on α.

---

[2]Use $\tilde{e}_n^{a,x}(T) := \inf\{a^k \,\mathrm{diam}(T) : \prod_{i=0}^k N(T\cap B(x, a^{i-2}\,\mathrm{diam}(T)), a^i\,\mathrm{diam}(T)) < 2^{2^n}\}$, where $N(A,\varepsilon)$ is the covering number of A by balls of radius ε. Details are left to the reader.

PROOF. As $0 \le K(t,x) \le f(x)$ and

$$K(t,x) - K(s,x) \ge (t-s)d\big(x, \pi_t(x)\big)$$

for every t, s, x, we have

$$\sum_{n\ge 0} a2^{n/\alpha} d\big(x, \pi_{a2^{n/\alpha}}(x)\big) \lesssim \sum_{n\ge 0} \big\{K\big(a2^{n/\alpha}, x\big) - K\big(a2^{(n-1)/\alpha}, x\big)\big\} \le f(x)$$

for every $x \in T$ and $a > 0$. □

Lemma 4.5 provides a natural mechanism to create multiscale approximations. However, the approximating sets $K_{a2^{n/\alpha}}$ are still continuous, and must therefore be discretized in order to bound the chaining functional that appears in Theorem 1.1. The simplest possible way to do this is to distribute each net T_n in Theorem 1.1 uniformly over the interpolation set $K_{a2^{n/\alpha}}$. This yields the basic interpolation bound of [12].

THEOREM 4.6. *For every $a > 0$, we have*

$$\gamma_\alpha(T) \lesssim \frac{1}{a} \sup_{x\in T} f(x) + \sum_{n\ge 0} 2^{n/\alpha} e_n(K_{a2^{n/\alpha}}),$$

where the universal constant depends only on α.

PROOF. By definition of entropy numbers, we can choose a set T_n with $|T_n| < 2^{2^n}$ such that $d(x, T_n) \le 2e_n(K_{a2^{n/\alpha}})$ for every $x \in K_{a2^{n/\alpha}}$. Then

$$e_n(A) \le \sup_{x\in A} d(x, T_n) \le \sup_{x\in A} d(x, K_{a2^{n/\alpha}}) + 2e_n(K_{a2^{n/\alpha}})$$

for every $A \subseteq T$. We can therefore invoke Theorem 3.1 with $a = 0$ and $s_n(x) = d(x, K_{a2^{n/\alpha}}) + 2e_n(K_{a2^{n/\alpha}})$. Lemma 4.5 completes the proof. □

The utility of Theorem 4.6 stems from the fact that the sets K_t are often much smaller than the index set T, so that this result provides a major improvement over Dudley's bound. This phenomenon is illustrated in various examples in [12]. Nonetheless, there is no reason to expect that the particular multiscale construction used here should always attain the sharp bound that is guaranteed by Theorem 1.1. Indeed, it is shown in [12], Section 3.3, that this is not necessarily the case.

There are two potential ways in which Theorem 4.6 can result in a suboptimal bound. First, the ability of this method to produce sufficiently "thin" sets K_t relies on a good choice of the penalty function f. While certain natural choices are considered in [12], the best choice is not always obvious, and a poor choice of penalty will certainly give rise to suboptimal bounds. This is, however, not an intrinsic deficiency of the interpolation method.

The fundamental inefficiency of Theorem 4.6 lies in the discretization of the sets K_t. The interpolation method cannot itself produce discrete nets: it only reveals a multiscale structure inside the index set T. To obtain the above result, we naively discretized this structure by distributing nets T_n as uniformly as possible over the sets $K_{a2^{n/\alpha}}$. While this provides an improvement over Dudley's bound, such a uniform discretization can incur a significant loss. In general, we should allow once again for a multiscale discretization of the sets $K_{a2^{n/\alpha}}$. It is easy to modify the above argument to formalize this idea; for example, one can easily show that

$$\gamma_\alpha(T) \lesssim \frac{1}{a} \sup_{x\in T} f(x) + \inf \sup_{x\in T} \sum_{n\geq 0} 2^{n/\alpha} d\big(\pi_{a2^{n/\alpha}}(x), T_n\big),$$

where inf is taken over all nets T_n with $|T_n| < 2^{2^n}$. This bound appears to be rather useless, however, as the quantity on the right-hand side is just as intractable as the quantity $\gamma_\alpha(T)$ that we wish to control in the first place.

The basic insight that gave rise to the results in this paper is that it is not actually necessary to construct explicit nets T_n to bound the right-hand side of this inequality: it suffices to show that the quantity on the right-hand side is significantly smaller than $\gamma_\alpha(T)$. For example, if we could show that

$$\inf \sup_{x\in T} \sum_{n\geq 0} 2^{n/\alpha} d\big(\pi_{a2^{n/\alpha}}(x), T_n\big) \lesssim a\gamma_\alpha(T),$$

then the resulting bound $\gamma_\alpha(T) \lesssim a\gamma_\alpha(T) + \frac{1}{a}\sup_{x\in T} f(x)$ would yield an explicit bound on $\gamma_\alpha(T)$ by choosing a to be sufficiently small. Such a bound captures quantitatively the idea that the sets K_t are much smaller than the index set T. The author initially implemented this idea in a special case (Section 5.1) using the formulation described above. It turns out, however, that the same scheme of proof is applicable far beyond this specific setting and is in some sense canonical. The contraction principle of Theorem 3.1 is nothing other than an abstract formulation of this idea that will enable us to efficiently exploit the interpolation method.

For future reference, we conclude this section by recording a convenient observation: the mapping $x \mapsto \pi_t(x)$ can often be chosen to be a (nonlinear) projection. This was established in [12] in a more restrictive setting.

LEMMA 4.7. *Suppose $T = \{x \in X : f(x) \leq u\}$. Then $K_t \subseteq T$ for all $t \geq 0$, and the minimizers $\pi_t(x)$ may be chosen to satisfy $\pi_t(\pi_t(x)) = \pi_t(x)$.*

PROOF. As $f(\pi_t(x)) \leq K(t,x) \leq f(x)$, we clearly have $\pi_t(x) \in T$ whenever $x \in T$. This shows that $K_t \subseteq T$. Now consider the set

$$K'_t := \big\{x \in T : K(t,x) = f(x)\big\}.$$

By construction, if $x \in K_t'$, we may choose $\pi_t(x) = x$. If $x \notin K_t'$, we choose $\pi_t(x)$ to be an arbitrary minimizer. We claim that $\pi_t(x) \in K_t'$ for all $x \in T$.

Indeed, suppose $\pi_t(x) \notin K_t'$. Then there exists $z \in X$ such that

$$f(z) + td\big(\pi_t(x), z\big) < f\big(\pi_t(x)\big).$$

But then we have

$$\begin{aligned} K(t,x) &= f\big(\pi_t(x)\big) + td\big(x, \pi_t(x)\big) \\ &> f(z) + td\big(\pi_t(x), z\big) + td\big(x, \pi_t(x)\big) \\ &\geq f(z) + td(x, z) \end{aligned}$$

by the triangle inequality. This contradicts the definition of $K(t,x)$.

As $\pi_t(x) \in K_t'$ for all $x \in T$, we have $K_t \subseteq K_t'$. On the other hand, as $x = \pi_t(x)$ for all $x \in K_t'$, it follows that $K_t = K_t'$ and $\pi_t(\pi_t(x)) = \pi_t(x)$. $\square$

5. Banach lattices and uniform convexity. In this section, we encounter our first nontrivial application of the contraction principle. We begin by developing in Section 5.1 a sharper version of a geometric principle that was obtained in [12], resolving a question posed in [12], Remark 4.4. We will use this principle in Section 5.2 to obtain a rather general geometric understanding of the behavior of the chaining functionals γ_α on Banach lattices. In Section 5.3, we discuss an analogous result for uniformly convex bodies.

5.1. *A geometric principle.* Throughout this section, we specialize our general setting to the case that $(X, \|\cdot\|)$ is a Banach space and $T \subset X$ is a symmetric compact convex set. We let $d(x,y) := \|x - y\|$, and denote the gauge of T as $\|x\|_T := \inf\{s \geq 0 : x \in sT\}$. It is natural in the present setting to use a power of the gauge as a penalty function in the interpolation method: that is, we define throughout this section

$$K(t,x) := \inf_{y \in X} \big\{\|y\|_T^r + t\|x - y\|\big\}$$

for some $r > 0$. The existence of minimizers $\pi_t(x)$ for $x \in T$ is easily established,[3] and we define as in Section 4.3 the interpolation sets

$$K_t := \big\{\pi_t(x) : x \in T\big\}.$$

We would like to impose geometric assumptions on the sets K_t that will allow us to obtain tractable bounds on $\gamma_\alpha(T)$. To this end, we will prove a sharper form of a useful geometric principle identified in [12], Theorem 4.1.

[3] As $K(t,x) \leq \|x\|_T^r \leq 1$, we may restrict the inf to be taken over the compact set $y \in T$. But $\|y\|_T = \sup_{z \in T^\circ} \langle z, y\rangle$, so the gauge is lower-semicontinuous and inf is attained.

THEOREM 5.1. *Let $q \geq 1$ and $L > 0$ be given constants, and suppose*

$$\|y - z\|_T^q \leq Lt\|y - z\| \qquad \text{for all } y, z \in K_t, t \geq 0.$$

Then

$$\gamma_\alpha(T) \lesssim \begin{cases} L^{1/q}\left[\sum_{n\geq 0}(2^{n/\alpha}e_n(T))^{q/(q-1)}\right]^{(q-1)/q} & (q > 1), \\ L\sup_{n\geq 0} 2^{n/\alpha}e_n(T) & (q = 1), \end{cases}$$

where the universal constant depends only on α.

The message of this result is that one can improve substantially on Dudley's inequality (which is the case $q = \infty$) if the geometric condition of Theorem 5.1 is satisfied. This condition is one manifestation of the idea that the sets K_t are much smaller than T: under this condition, every small ball in K_t is contained in a proportionally scaled-down copy of T. Of course, it is not at all obvious how to realize this condition, but we will see below that it arises very naturally from the interpolation method under suitable geometric assumptions on T.

For fixed $q > 1$, it was shown in [12], Theorem 4.1, that the conclusion of Theorem 5.1 can be deduced from Theorem 4.6. However, this approach has a crucial drawback: the constant diverges as $q \downarrow 1$. The key improvement provided by Theorem 5.1 is that the constant does not depend on q, which allows us in particular to attain the limiting case $q = 1$. The latter is particularly interesting, as the so-called Sudakov lower bound

$$\gamma_\alpha(T) \geq \sup_{n\geq 0} 2^{n/\alpha}e_n(T)$$

holds trivially for any T. Thus the case $q = 1$ of Theorem 5.1 gives a geometric condition for the Sudakov lower bound to be sharp, as conjectured in [12], Remark 4.4. We will encounter an important example in Section 5.2.

PROOF OF THEOREM 5.1. Let $n \geq 0$ and $A \subseteq T$. We denote by

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

$$s(t, A) := \sup_{x\in A}\|x - \pi_t(x)\|$$

the projection of A on K_t and the associated projection error.

We first note that the assumption of the theorem implies that

$$A_t \subseteq z + (Lt\,\mathrm{diam}(A_t))^{1/q}T$$

for some $z \in X$. That is, the projection A_t is contained in a "shrunk" copy of T. On the other hand, replacing A_t by A only costs the projection error:

$$e_n(A) \leq e_n(A_t) + s(t, A),$$

$$\mathrm{diam}(A_t) \leq \mathrm{diam}(A) + 2s(t, A).$$

We can therefore estimate

$$e_n(A) \le (Lt)^{1/q}\big(\operatorname{diam}(A) + 2s(t, A)\big)^{1/q} e_n(T) + s(t, A).$$

We apply this bound with $t = a2^{n/\alpha}$. The idea is now that the interpolation lemma will take care of the projection error, while the "contraction" part of the contraction principle allows us to exploit the shrinkage created by the geometric assumption.

Case $q = 1$. In this case, we can estimate

$$\begin{aligned} e_n(A) &\le LSa\operatorname{diam}(A) + (2LSa + 1)s\big(a2^{n/\alpha}, A\big), \\ S &:= \sup_{n\ge 0} 2^{n/\alpha} e_n(T). \end{aligned}$$

Applying the contraction principle of Theorem 3.1 gives

$$\gamma_\alpha(T) \lesssim LSa\gamma_\alpha(T) + (2LSa + 1)\sup_{x\in T}\sum_{n\ge 0} 2^{n/\alpha}\big\|x - \pi_{a2^{n/\alpha}}(x)\big\|.$$

We can now use the interpolation Lemma 4.5 to bound the second term as

$$\gamma_\alpha(T) \lesssim LSa\gamma_\alpha(T) + LS + \frac{1}{a}.$$

We conclude by setting $a = C/LS$ for a sufficiently small universal constant C.

Case $q > 1$. The proof is very similar, but now we use Young's inequality $uv \le u^p/C^{p/q} + Cv^q$ with $p = q/(q-1)$ to estimate

$$e_n(A) \le C\operatorname{diam}(A) + (2C + 1)s\big(a2^{n/\alpha}, A\big) + \left(\frac{La}{C}\right)^{p/q} 2^{np/\alpha q} e_n(T)^p.$$

If C is chosen to be a sufficiently small universal constant, then the contraction principle and interpolation lemma give, respectively,

$$\begin{aligned} \gamma_\alpha(T) &\lesssim \sup_{x\in T}\sum_{n\ge 0} 2^{n/\alpha}\big\|x - \pi_{a2^{n/\alpha}}(x)\big\| + (La)^{p/q}\sum_{n\ge 0}\big(2^{n/\alpha} e_n(T)\big)^p \\ &\lesssim \frac{1}{a} + (La)^{p/q}\sum_{n\ge 0}\big(2^{n/\alpha} e_n(T)\big)^p \end{aligned}$$

(here we used in the first line that $2^{n/\alpha}2^{np/\alpha q} = 2^{np/\alpha}$ because $1 + p/q = p$). The proof is completed by optimizing over $a > 0$. □

Let us note that the choice of $r > 0$ in the definition of $K(t, x)$ appears nowhere in the statement of proof of Theorem 5.1. The flexibility to choose r in a convenient manner will be useful, however, when we try to verify that the assumption of Theorem 5.1 is satisfied in specific situations.

5.2. *Banach lattices*. The aim of this section is to show that Theorem 5.1 provides a rather general understanding of the behavior of $\gamma_\alpha(T)$ on Banach lattices (all the relevant background on Banach lattices and their geometry can be found in [5]). To this end, we specialize the setting of the previous section to the case where $(X, \|\cdot\|)$ is a Banach lattice and where the compact convex set $T \subset X$ is solid, that is, $x \in T$ and $|y| \le |x|$ implies $y \in T$. Solidity of T is simply the requirement that the gauge $\|\cdot\|_T$ is also a lattice norm (on its domain). We now introduce a fundamental property that plays an important role in the geometry of Banach lattices; cf. [5], Section 1.f.

DEFINITION 5.2. Let $q \ge 1$. T is said to satisfy a *lower q-estimate* with constant M if for all $n \ge 1$ and vectors $x_1, \dots, x_n \in X$

$$\left[\sum_{i=1}^n \|x_i\|_T^q\right]^{1/q} \le M \left\|\sum_{i=1}^n |x_i|\right\|_T.$$

We have the following result.

THEOREM 5.3. *If T satisfies a lower q-estimate with constant M, then*

$$\gamma_\alpha(T) \lesssim \begin{cases} M\left[\sum_{n\ge 0}(2^{n/\alpha}e_n(T))^{q/(q-1)}\right]^{(q-1)/q} & (q>1), \\ M \sup_{n\ge 0} 2^{n/\alpha} e_n(T) & (q=1), \end{cases}$$

where the universal constant depends only on α.

We will prove this theorem by showing that the condition of Theorem 5.1 is satisfied if we choose $r = q$ in the previous section. There is a somewhat subtle point, however, that we must take care of first. The computations used in our proof rely crucially on the fact that a lower q-estimate is satisfied with constant $M = 1$. However, we did not require this special situation to hold in Theorem 5.3. We will therefore make essential use of the observation that any Banach lattice that satisfies a lower q-estimate admits an equivalent renorming whose lower q-estimate constant is identically one ([5], Lemma 1.f.11). Concretely, define the new norm

$$\|x\|_{\tilde{T}} := \sup\left[\sum_{i=1}^n \|x_i\|_T^q\right]^{1/q},$$

where the supremum is taken over all possible decompositions of x as a sum of $n \ge 1$ pairwise disjoint elements $x_1, \dots, x_n$, and define $\tilde{T} := \{x \in X : \|x\|_{\tilde{T}} \le 1\}$. It is readily verified using [5], Proposition 1.f.6, that if T satisfies a lower q-estimate with constant M, then $\tilde{T}$ satisfies a lower q-estimate with constant 1 and $\tilde{T} \subseteq T \subseteq M\tilde{T}$. This implies in particular that $\gamma_\alpha(T) \le M\gamma_\alpha(\tilde{T})$ and $e_n(\tilde{T}) \le e_n(T)$, so that we may assume without loss of generality in the proof of Theorem 5.3 that $M = 1$.

PROOF OF THEOREM 5.3. We will assume without loss of generality that $M = 1$, and we apply the setting of the previous section with $r = q$. Fix $t \ge 0$ and $y, z \in K_t$, and define

$$u := (y \wedge z) \vee 0 + (y \vee z) \wedge 0.$$

The point of this definition is that

$$|y| - |u| = |y - u| \le |y - z|,$$

as well as the analogous property where the roles of y and z are exchanged.

Using that T satisfies a lower q-estimate with constant one, we obtain

$$\|y - u\|_T^q \le \|y\|_T^q - \|u\|_T^q.$$

On the other hand, Lemma 4.7 gives

$$\|y\|_T^q = K(t, y) \le \|u\|_T^q + t\|y - u\|.$$

All the above properties hold if we exchange y and z. Therefore,

$$\begin{aligned}\|y - z\|_T^q &\le 2^{q-1}\big(\|y - u\|_T^q + \|z - u\|_T^q\big)\\ &\le 2^{q-1}t\big(\|y - u\| + \|z - u\|\big)\\ &\le 2^q t\|y - z\|,\end{aligned}$$

where we used the triangle inequality, $(a + b)^q \le 2^{q-1}(a^q + b^q)$, and that $\|\cdot\|$ is a lattice norm. The proof is concluded by applying Theorem 5.1. □

An interesting example of Theorem 5.3 is the the following. Let $X = \mathbb{R}^d$, let $\|\cdot\|$ be any 1-unconditional norm (with respect to the standard basis), and let $T = B_1^d$ be the unit ℓ_1-ball. It is immediate that the ℓ_1-norm satisfies a lower 1-estimate with constant one. Theorem 5.3 therefore yields

$$\gamma_\alpha\big(B_1^d\big) \asymp \sup_{n\ge 0} 2^{n/\alpha} e_n\big(B_1^d\big),$$

that is, Sudakov's lower bound is sharp for the ℓ_1-ball. In the special case where $\alpha = 2$ and $\|\cdot\|$ is the Euclidean norm, this can be verified by an explicit computation using Theorem 1.1 (or Theorem 4.6, cf. [12], Section 3.2) and estimates on the entropy numbers; however, such a computation does not explain *why* Sudakov turns out to be sharp in this setting. Theorem 5.3 provides a geometric explanation of this phenomenon, and extends it to the much more general situation where $\|\cdot\|$ is an arbitrary unconditional norm.

We conclude this section with a few remarks.

REMARK 5.4. We have shown that Sudakov's inequality is sharp for B_1^d if $\|\cdot\|$ is a lattice norm (i.e., unconditional with respect to the standard basis). It is worth noting that the lattice property is really essential for this phenomenon to occur: the analogous result for general norms is absolutely false. To see why this must be the case, note that if T is the symmetric convex hull of d points in X, then we always have $T = AB_1^d$ for some linear operator $A : \mathbb{R}^d \to X$. We can therefore write $\gamma_\alpha(T, \|\cdot\|) = \gamma_\alpha(B_1^d, \|\cdot\|')$ with $\|x\|' := \|Ax\|$. Thus if Sudakov's lower bound were sharp for B_1^d when endowed with a general norm, then Sudakov's lower bound would be sharp for any symmetric polytope and, therefore, (by approximation) for every symmetric compact convex set. This conclusion is clearly false.

REMARK 5.5. The case $q = 1$ of Theorem 5.3 proves to be somewhat restrictive. Suppose that $\|\cdot\|_T$ satisfies a lower 1-estimate with constant one (as may always be assumed after equivalent renorming). Because of the triangle inequality, we must then have the rather strong condition $\|x\|_T + \|y\|_T = \|(|x| + |y|)\|_T$. A Banach lattice satisfying this condition is called an AL-space. It was shown by Kakutani that such a space is always order-isometric to $L^1(\mu)$ for some measure (μ [5], Theorem 1.b.2). Thus L^1-balls are essentially the only examples for which Theorem 5.3 applies with $q = 1$. The case $q > 1$ is much richer, however, and Theorem 5.3 provides a very general tool to understand chaining functionals in this setting.

REMARK 5.6. Theorem 5.3 shows that Dudley's inequality can be substantially improved for solid sets T that satisfy a nontrivial lower q-estimate. On the other hand, a solid set T that fails to satisfy any nontrivial lower q-estimate must contain ℓ_∞^d-balls of arbitrarily large dimension; cf. [5], Theorem 1.f.12. For cubes, the majorizing measure theorem and the results of [2] can be used to show that Dudley's inequality is sharp, and that no improvement as in Theorem 5.3 can hold in general. Thus Theorem 5.3 is essentially the best result of its kind.

5.3. *Uniformly convex bodies*. The lower q-estimate property of a Banach lattice is closely related to the notion of uniform convexity in general Banach spaces, as is explained in [5], Section 1.f. It is therefore not surprising that an analogue of Theorem 5.3 holds in a general Banach space when T is a uniformly convex body. Unlike the results of the previous section, however, this case has already been well understood ([11], Section 4.1). It will nonetheless be useful to revisit this setting in the light of the present paper, as the method that appears in the proof will play an essential role in the random matrix problems that will be discussed in Section 7.

To this end, we return to the setting where $(X, \|\cdot\|)$ is a general Banach space and $T \subset X$ is a symmetric compact convex set.

DEFINITION 5.7. Let $q \geq 2$. T is said to be *q-convex* with constant η if

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

for all vectors $x, y \in T$.

It was shown in [12], Lemma 4.7, that the assumption of Theorem 5.1 holds in the present setting with $L = 1/2\eta$; the proof of this fact is not unlike the one we used in the lattice case. Thus the conclusion in the case q-convex bodies matches verbatim the one obtained for lattices in the previous section. However, in this setting we are never near the boundary case of Theorem 5.1, as the q-convexity property can only hold for $q \geq 2$ (no body is more strongly convex than a Euclidean ball). This means that the machinery of this paper is not really needed to establish this result; it was shown in [12] that the same conclusion already follows from Theorem 4.6.

However, the boundary case reappears if we consider the more general chaining functionals $\gamma_{\alpha,p}(T)$ rather than $\gamma_\alpha(T)$. For example, the following result of [11], Theorem 4.1.4, cannot be recovered by the methods of [12].

THEOREM 5.8. *Let $q \geq 2$. If T is q-convex with constant η, then*

$$\gamma_{\alpha,q}(T) \lesssim \eta^{-1/q} \sup_{n\geq 0} 2^{n/\alpha} e_n(T),$$

where the universal constant depends only on α.

We will presently give a short proof of this result using the methods of this paper. Of course, one can obtain extensions of both Theorems 5.3 and 5.8 that bound $\gamma_{\alpha,p}(T)$ with general $\alpha > 0$ and $1 \leq p \leq q$ (not just $p = q$ as in Theorem 5.8); as no new ideas arise in this setting, we omit the details.

To bound $\gamma_{\alpha,q}$, we require in principle only a minor adaptation of the interpolation method: we modify the definition of $K(t,x)$ in Section 5.1 to

$$K(t,x) := \inf_{y\in X} \{\|y\|_T^r + t^q \|x-y\|^q\}.$$

Denote once again by $\pi_t(x)$ the minimizer in this expression, and by K_t the set of minimizers for $x \in T$. The analogue of the interpolation lemma in this setting is obtained by repeating verbatim the proof of Lemma 4.5.

LEMMA 5.9. *For every $a > 0$, we have*

$$\sup_{x\in T} \sum_{n\geq 0} (2^{n/\alpha} \|x - \pi_{a2^{n/\alpha}}(x)\|)^q \lesssim \frac{1}{a^q},$$

where the universal constant depends only on α.

With these simple modifications, we can now essentially follow the same scheme of proof as for Theorem 5.3, replacing the use of the lower q-estimate by the q-convexity property. There is, however, one minor issue that requires some care. In the proof of Theorem 5.3 (as in the proof of [12], Lemma 4.7, where the assumption of Theorem 5.1 is verified for q-convex sets), we used the fact that $\pi_t(x)$ possesses the projection property of Lemma 4.7. This property is however quite special to interpolation functionals of the form $\inf\{f(y)+td(x,y)\}$, as it relies crucially on the triangle property of the distance. When the distance is raised to a power as in the present setting, the projection property no longer holds and we must take care to proceed without it. Fortunately, it turns out to that the projection property was not really used in an essential way in Theorem 5.3 and can easily be avoided.

PROOF OF THEOREM 5.8. Applying Definition 5.7 to x/γ, y/γ with $\gamma = \max(\|x\|_T, \|y\|_T) \le 1$, the q-convexity property can be formulated as

$$\left\|\frac{x+y}{2}\right\|_T \le \max\bigl(\|x\|_T, \|y\|_T\bigr) - \eta\|x-y\|_T^q$$

for every $x, y \in T$. To exploit this formulation of q-convexity, we will choose $r = 1$ in the definition of $K(t,x)$.

Let $n \ge 0$ and $A \subseteq T$. As in the proof of Theorem 5.1, we write

$$A_t := \bigl\{\pi_t(x) : x \in A\bigr\},$$
$$s(t,A) := \sup_{x\in A}\bigl\|x - \pi_t(x)\bigr\|.$$

Note that $A_t \subseteq T$. If $y = \pi_t(x)$ for $x \in A$, we can estimate

$$\begin{aligned}
\|y\|_T &\le \|y\|_T + t^q\|x-y\|^q = K(t,x)\\
&\le \|u\|_T + t^q\|x-u\|^q\\
&\le \|u\|_T + 2^{q-1}t^q\|y-u\|^q + 2^{q-1}t^q\|x-y\|^q\\
&\le \|u\|_T + 2^{q-1}t^q\|y-u\|^q + 2^{q-1}t^q s(t,A)^q
\end{aligned}$$

for any $u \in X$, where we used the triangle inequality and $(a+b)^q \le 2^{q-1}(a^q+b^q)$. Thus for any $y, z \in A_t$, choosing $u := (y+z)/2$ yields

$$\max\bigl(\|y\|_T, \|z\|_T\bigr) \le \left\|\frac{y+z}{2}\right\|_T + 2^{q-1}t^q\left\|\frac{y-z}{2}\right\|^q + 2^{q-1}t^q s(t,A)^q.$$

Applying the q-convexity property yields

$$\eta\|y-z\|_T^q \le 2^{-1}t^q\|y-z\|^q + 2^{q-1}t^q s(t,A)^q$$

for every $y, z \in A_t$. Note that this condition is almost identical to the assumption of Theorem 5.1, except that an additional projection error term appears. The latter is the price we pay for avoiding the projection property, which does not hold in the present setting. However, this additional term introduces no further complications in the proof, and we can simply proceed as before. The above inequality shows that

$$A_t \subseteq z + \eta^{-1/q} t\big(\mathrm{diam}(A_t) + 2s(t, A)\big)T$$

for some $z \in X$. Proceeding as in the proof of Theorem 5.1, we obtain

$$\begin{aligned} e_n(A) &\le Sa\,\mathrm{diam}(A) + (4Sa+1)s\big(a2^{n/\alpha}, A\big), \\ S &:= \eta^{-1/q} \sup_{n\ge 0} 2^{n/\alpha} e_n(T). \end{aligned}$$

Applying Theorem 3.1 and Lemma 5.9 yields

$$\begin{aligned} \gamma_{\alpha,q}(T) &\lesssim Sa\gamma_{\alpha,q}(T) + (4Sa+1)\left[\sup_{x\in T}\sum_{n\ge 0}\big(2^{n/\alpha}\|x - \pi_{a2^{n/\alpha}}(x)\|\big)^q\right]^{1/q} \\ &\lesssim Sa\gamma_{\alpha,q}(T) + S + \frac{1}{a}. \end{aligned}$$

We now choose $a = C/S$ for a sufficiently small universal constant C. $\square$

It is also possible to give a proof more in the spirit of Theorem 5.3 where we choose $r = q$ in the definition of $K(t, x)$. In this case, one should replace Definition 5.7 by the following homogeneous form of q-convexity:

$$\left\|\frac{x+y}{2}\right\|_T^q \le \frac{\|x\|_T^q + \|y\|_T^q}{2} - \tilde{\eta}\|x-y\|_T^q$$

for all $x, y \in X$. It can be shown that this alternative formulation is equivalent to that of Definition 5.7 [1], Proposition 7, and a more careful accounting of the constants (as in [6], Lemma 2.2) shows that $\tilde{\eta} \ge c^q \eta$.

REMARK 5.10. As was mentioned above, the analogue of Theorem 5.3 for q-convex sets was already proved in [12] using only Theorem 4.6: one can show in this case that the entropy numbers of the interpolation sets $e_n(K_t)$ can be controlled efficiently by the entropy numbers $e_n(T)$. It was even shown in [12] by explicit computation that Theorem 4.6 yields a sharp bound for the ℓ_1-ball in the special case that $\|\cdot\|$ is the Euclidean norm, which is a boundary case of Theorem 5.3. This is simpler conceptually than the present approach, which relies on the contraction principle. One might therefore wonder whether the contraction principle is truly needed in this setting, or whether results such as Theorems 5.3 and 5.8 could be recovered from Theorem 4.6 using a more efficient argument. We will presently argue that this is not the case: the entropy numbers $e_n(K_t)$ are generally too large, so the contraction principle is essential to attain sharp bounds.

To this end, consider the following example. Consider $X = \mathbb{R}^d$ with the Euclidean distance $\|\cdot\|$, and let $T \subset X$ be the ellipsoid defined by

$$\|x\|_T^2 = \sum_{k=1}^d k x_k^2.$$

T is 2-convex by the parallelogram identity, and Theorem 5.8 gives

$$\gamma_{2,2}(T) \asymp \sup_{n\geq 0} 2^{n/2} e_n(T) \asymp 1$$

as $e_n(T) \lesssim 2^{-n/2}$ by the estimates in [11], Section 2.5.

It is trivial to adapt Theorem 4.6 the present setting, which yields

$$\gamma_{2,2}(T) \lesssim \frac{1}{a} + \left[\sum_{n\geq 0} \left(2^{n/2} e_n(K_{a2^{n/2}})\right)^2\right]^{1/2} =: S(a).$$

We claim that this bound cannot recover the correct behavior of $\gamma_{2,2}(T)$. To see this, we must compute the interpolation sets K_t. It is particularly convenient in this setting to choose $r = 2$ in the definition of $K(t, x)$, which is appropriate as explained after the proof of Theorem 5.8. The advantage of this choice is that $K(t, x) := \inf_y \{\|y\|_T^2 + t^2\|x - y\|^2\}$ involves minimizing a quadratic function. We readily compute that K_t is another ellipsoid:

$$\left(\pi_t(x)\right)_k = \frac{t^2}{t^2 + k} x_k,$$

$$\|x\|_{K_t} = \sum_{k=1}^d \left(\frac{t^2 + k}{t^2}\right)^2 k x_k^2.$$

Using the entropy estimate of [11], Lemma 2.5.4, we find that

$$e_n(K_{a2^{n/2}}) \gtrsim \frac{a^2}{a^2 + 1} 2^{-n/2}$$

for $2^n \lesssim d$. It follows that

$$\begin{aligned} S(a) &= \frac{1}{a} + \left[\sum_{n\geq 0} \left(2^{n/2} e_n(K_{a2^{n/2}})\right)^2\right]^{1/2} \\ &\gtrsim \frac{1}{a} + \frac{a^2}{a^2 + 1}\sqrt{\log d} \\ &\gtrsim (\log d)^{1/6}. \end{aligned}$$

We have therefore shown that a sharp bound on $\gamma_{2,2}(T)$ cannot be attained by choosing nets that are distributed uniformly on the interpolation sets K_t, as is done in Theorem 4.6. On the other hand, precisely the same interpolation scheme yields a sharp bound when combined with the contraction principle in Theorem 5.8. This example provides an explicit illustration of the assertion made in the Introduction that the deficiency of Theorem 4.6 is not due to the interpolation method, but rather due to the fact that the interpolation method is being used inefficiently.

6. The majorizing measure theorem. In the previous sections we introduced the contraction principle and illustrated its utility in combination with the interpolation method in several situations. We will presently use the same machinery to give a surprisingly simple proof of the majorizing measure theorem (Theorem 1.1). With some small modifications, this will also allow us to recover the main growth functional estimate of [11]. Beside providing simple new proofs of these results, the fact that they can be attained at all shows that the methods of this paper are not restricted to some special situations, but can in fact recover the core of the generic chaining theory.

6.1. *Gaussian processes.* Let $(X_x)_{x\in T}$ be a centered Gaussian process, and let $d(x,y) := (\mathbf{E}|X_x - X_y|^2)^{1/2}$ be the associated natural metric on T. To avoid being distracted by minor measurability issues, we will assume for simplicity that the index set T is finite. It is well understood in the theory of Gaussian processes that this entails no loss of generality.

Let us define for any subset $A \subseteq T$ the Gaussian width

$$G(A) := \mathbf{E}\Big[\sup_{x\in A} X_x\Big].$$

The statement of the majorizing measure theorem is that $G(T) \asymp \gamma_2(T)$. The upper bound $G(T) \lesssim \gamma_2(T)$ is however completely elementary; see [11], Section 2.2, for this classical and very simple chaining argument. It is the lower bound $\gamma_2(T) \lesssim G(T)$ in the majorizing measure theorem that is a deep result. In this section, we will give a very simple proof of the latter bound using the machinery of this paper.

In its simplest form, the idea that allows us to bound $\gamma_2(T)$ by $G(T)$ is clear: we should use $G(T)$ to define the penalty function in the interpolation method, and then use Sudakov's inequality for Gaussian processes [which gives an upper bound on $e_n(A)$ in terms of $G(A)$] to verify the assumption of the contraction principle. To implement this idea, it will be convenient to define the interpolation functional $K(t,x)$ in a somewhat different manner than we did previously: we define throughout this section

$$K(t,x) := \inf_{s\geq 0}\big\{ts + G(T) - G\big(B(x,s)\big)\big\},$$

where

$$B(x,s) := \big\{y \in T : d(x,y) \leq s\big\}$$

is the ball in T with radius s centered at x. As $s \mapsto G(B(x,s))$ is upper-semicontinuous, the infimum in the definition of $K(t,x)$ is attained. Denoting the minimizer as $s(t,x) \geq 0$, we obtain the following interpolation lemma.

LEMMA 6.1. *For every $a > 0$,*

$$\sup_{x\in T}\sum_{n\geq 0} 2^{n/2} s\big(a2^{n/2}, x\big) \lesssim \frac{G(T)}{a}.$$

PROOF. The proof is identical to that of Lemma 4.5, but we include it for completeness. Note that, by definition of $K(t,x)$,

$$\begin{aligned} K(t,x) - K(u,x) &= ts(t,x) - G\big(B\big(x, s(t,x)\big)\big) - \inf_{s\geq 0}\big\{us - G\big(B(x,s)\big)\big\} \\ &\geq (t-u)s(t,x) \end{aligned}$$

[choose $s = s(t,x)$ in the infimum on the right], and that $0 \leq K(t,x) \leq G(T)$ [choose $s = 0$ in the definition of $K(t,x)$]. Therefore,

$$\sum_{n\geq 0} a2^{n/2} s\big(a2^{n/2}, x\big) \lesssim \sum_{n\geq 0}\big\{K\big(a2^{n/2}, x\big) - K\big(a2^{(n-1)/2}, x\big)\big\} \leq G(T),$$

and the proof is complete. □

REMARK 6.2. It may not be obvious that the present definition of $K(t,x)$ is an interpolation functional in the sense of Section 4.3, except in some generalized sense. This is nonetheless the case. To see why, let $X := L^\infty(\Omega; T)$ be the space of T-valued bounded random variables. We identify $T \subset X$ with the set of constant functions, and extend the natural metric on T to X according to $d(\sigma,\tau) := \sup_{\omega\in\Omega} d(\sigma(\omega), \tau(\omega))$. Then we have for any $x \in T$

$$G\big(B(x,s)\big) = \mathbf{E}\Big[\sup_{y\in T: d(x,y)\leq s} X_y\Big] = \sup_{\tau\in X: d(x,\tau)\leq s} \mathbf{E}[X_\tau].$$

Substituting this expression in the definition of $K(t,x)$ and exchanging the order of the two infima shows that we can in fact write

$$K(t,x) = \inf_{\tau\in X}\big\{G(T) - \mathbf{E}[X_\tau] + td(x,\tau)\big\}.$$

Thus $K(t,x)$ is an interpolation functional, on the space (X,d) with penalty function $f(\tau) = G(T) - \mathbf{E}[X_\tau]$, of precisely the form given in Section 4.3. While this formulation guides our intuition, it is more convenient computationally to work with the definition in terms of $G(B(x,s))$ as this will allow us to directly apply inequalities for the suprema of Gaussian processes.

To prove the majorizing measure theorem, we will verify that the condition of Theorem 3.1 is satisfied with $s_n(x) \lesssim s(a2^{n/2}, x)$. To this end, we must bound the entropy numbers $e_n(A)$ of all subsets $A \subseteq T$. As our interpolation functional involves the supremum of a Gaussian process, this should surely involve Sudakov's inequality. The appropriate form for our purposes, which is a straightforward combination of Sudakov's inequality and Gaussian concentration, can be found in [11], Proposition 2.4.9.

LEMMA 6.3. *For $\sigma, b > 0$, $x_1, \dots, x_n \in T$ so that $d(x_i, x_j) \geq b$ for $i \neq j$*

$$\min_{i \leq n} G\big(B(x_i, \sigma)\big) + C_1 b\sqrt{\log n}$$
$$\leq G\left(\bigcup_{i \leq n} B(x_i, \sigma)\right) + C_2\sigma\sqrt{\log n},$$

where C_1, C_2 are universal constants.

This is in fact a form of the "growth condition" that forms the central ingredient in the generic chaining theory as developed in [11]. One of the advantages of the approach developed in this paper is that it makes it possible to bound chaining functionals without engineering such a condition, which does not always arise naturally in a geometric setting. However, in the case of Gaussian processes, the growth condition arises in a completely natural manner and is essentially the reason why the majorizing measure theorem is true. It therefore seems likely that any proof of the majorizing measure theorem must exploit a form of Lemma 6.3 at the crucial point in the argument. We will presently show that Lemma 6.3 provides a very simple method for verifying the assumption of the contraction principle.

LEMMA 6.4. *For every $n \geq 0$, $A \subseteq T$, and $a > 0$, we have*

$$e_n(A) \lesssim a \operatorname{diam}(A) + (a+1) \sup_{x \in A} s\big(a2^{n/2}, x\big).$$

PROOF. Assume $e_n(A) > 0$, else the result is trivial. By Lemma 2.2, find $N = 2^{2^n}$ points $x_1, \dots, x_N \in A$ so that $d(x_i, x_j) > e_n(A)/2$ for $i \neq j$. Let

$$\sigma = \sup_{x \in A} s\big(a2^{n/2}, x\big),$$
$$r = \operatorname{diam}(A) + \sigma.$$

Then $\bigcup_{i \leq N} B(x_i, \sigma) \subseteq B(x_k, r)$ for every $k \leq N$. We can now estimate

$$\begin{aligned} G(T) - G\big(B(x_k, \sigma)\big) &\leq G(T) - G\big(B\big(x_k, s\big(a2^{n/2}, x_k\big)\big)\big) \\ &\leq K\big(a2^{n/2}, x_k\big) \\ &\leq a2^{n/2} r + G(T) - G\big(B(x_k, r)\big) \\ &\leq a2^{n/2} r + G(T) - G\left(\bigcup_{i \leq N} B(x_i, \sigma)\right) \end{aligned}$$

for every $k \leq N$. Applying Lemma 6.3 gives

$$2^{n/2} e_n(A) \lesssim a2^{n/2} r + 2^{n/2}\sigma,$$

which readily yields the conclusion. □

With this lemma in hand, the proof of the lower bound in the majorizing measure theorem follows immediately from the contraction principle.

THEOREM 6.5. $\gamma_2(T) \lesssim G(T)$.

PROOF. The condition of Theorem 3.1 is verified by Lemma 6.4. It remains to apply Lemma 6.1 and choose $a > 0$ sufficiently small. $\square$

6.2. *Growth functionals*. Now that we have proved the majorizing measure theorem using the approach of this paper, it will come as no surprise that the general growth functional machinery that forms the foundation of the generic chaining theory as developed in [11] can also be recovered by the interpolation method. This shows that applicability of the interpolation method is not restricted to some special situations, but that it is in principle canonical: the generic chaining theory can be fully recovered in this manner. In our approach, growth functionals provide one possible method for creating the condition of the contraction principle.

In this section, we will modify the proof of the majorizing measure theorem to utilize one of the generalized growth functional conditions considered in [11]. While the basic idea of the proof is already contained in the previous section, this generalization is instructive in its own right: it will help clarify the relevance of the ingredients in the definition of a growth functional from the present perspective, and will also illustrate the use of different interpolation functionals for different scales. Of course, the same method of proof admits numerous generalizations, including several considered in [11] that can be analogously recovered by our methods.

We will work on a general metric space (T, d). Let us first state some definitions. The first is a notion of well-separated sets ([11], Definition 2.3.8).

DEFINITION 6.6. A collection of sets $H_1, \dots, H_m \subseteq T$ is said to be (b, c)-*separated* if there exist $x_1, \dots, x_m, y \in T$ such that $d(x_i, x_j) \geq b$ for all $i \neq j$, and $d(x_i, y) \leq cb$ and $H_i \subseteq B(x_i, b/c)$ for all i.

We also need the basic notion of a functional.

DEFINITION 6.7. A *functional* on T is a map F that assigns to every set $H \subseteq T$ a number $F(H) \geq 0$ and that is increasing, that is, $F(H) \leq F(H')$ if $H \subseteq H'$. A sequence of functionals $(F_n)_{n\geq 0}$ is said to be *decreasing* if $F_{n+1}(H) \leq F_n(H)$ for every set H.

We now state the growth condition of [11], Definition 2.3.10.

DEFINITION 6.8. A decreasing sequence of functionals $(F_n)_{n\geq 0}$ satisfies the *growth condition* with parameters $c, L > 0$ if for every $b > 0$, $n \geq 1$ and collection $H_1, \dots, H_N \subseteq T$ of $N = 2^{2^n}$ sets that is (b, c)-separated, we have

$$F_{n-1}\left(\bigcup_{i\leq N} H_i\right) \geq L2^{n/2}b + \min_{i\leq N} F_n(H_i).$$

A minor variation on Lemma 6.3 shows that the choice $F_n(H) = G(H)$ satisfies the growth condition provided the parameter c is chosen sufficiently large: that is, the Gaussian width $G(H)$ is a growth functional. However, the growth condition allows more flexibility in the design of functionals. Our aim is to prove the following result of [11], Theorem 2.3.16.

THEOREM 6.9. *Suppose the decreasing sequence of functionals $(F_n)_{n\geq 0}$ satisfies the growth condition with parameters $c, L > 0$. Then*

$$\gamma_2(T) \lesssim \frac{c}{L} F_0(T) + \mathrm{diam}(T)$$

provided that $c \geq c_0$, where c_0 is a universal constant.

In the sequel, we fix parameters $c, L > 0$ and a decreasing sequence of functionals $(F_n)_{n\geq 0}$ that satisfies the growth condition of Definition 6.2.

There are two additional ideas in the proof of Theorem 6.9 as compared to that of the majorizing measure theorem. First, we have not one growth functional G, but rather a separate functional F_n for every scale. This flexibility introduces makes the growth condition easier to satisfy. The complication that arises is that we have to work with multiple interpolation functionals

$$K_n(t, x) := \inf_{s\geq 0}\{ts + F_0(T) - F_n\big(B(x, s)\big)\}.$$

However, as F_n is a decreasing sequence of functionals, we readily recover a variant of the usual interpolation lemma. In the sequel, we will choose $s_n^a(x) \geq 0$ for every $n \geq 1$, $a > 0$ and $x \in T$ such that

$$\begin{aligned} K_n\big(La2^{n/2}, x\big) &\leq La2^{n/2}s_n^a(x) + F_0(T) - F_n\big(B\big(x, s_n^a(x)\big)\big) \\ &\leq K_n\big(La2^{n/2}, x\big) + 2^{-n}F_0(T). \end{aligned}$$

That is, $s_n^a(x)$ is a near-minimizer in the definition of $K_n(t, x)$. (As we assumed no regularity on F_n, we cannot guarantee that a true minimizer exists; but this makes no difference whatsoever in the proofs.)

LEMMA 6.10. *For every $a > 0$,*

$$\sup_{x\in T} \sum_{n\geq 1} 2^{n/2}s_n^a(x) \lesssim \frac{F_0(T)}{La}.$$

PROOF. By definition of K_{n-1} and as F_n is a decreasing sequence,

$$\begin{aligned}
&2^{-n}F_0(T)+K_n\big(La2^{n/2},x\big)-K_{n-1}\big(La2^{(n-1)/2},x\big)\\
&\quad\geq \big(1-2^{-1/2}\big)La2^{n/2}s_n^a(x)+F_{n-1}\big(B\big(x,s_n^a(x)\big)\big)-F_n\big(B\big(x,s_n^a(x)\big)\big)\\
&\quad\geq \big(1-2^{-1/2}\big)La2^{n/2}s_n^a(x).
\end{aligned}$$

We conclude by summing over $n\geq 1$ and using $K_n(t,x)\leq F_0(T)$. □

The second new feature in the proof of Theorem 6.9 is that the separation condition of Definition 6.6 is rather restrictive: it requires the sets H_i to have small diameter and all the points x_i to be close together. This provides, once again, more room in the growth condition of Definition 6.8 (as the growth condition must only hold for separated sets satisfying these restrictive assumptions). However, we will see in the proof of Lemma 6.11 below that these additional restrictions arise essentially for free: if either of these restrictions is violated, the condition of the contraction principle is automatically satisfied and there is nothing to prove.

LEMMA 6.11. *Fix $a>0$. Let $s_0(x):=\operatorname{diam}(T)$ and for $n\geq 1$*

$$s_n(x):=(a+c)s_n^a(x)+\frac{1}{L2^{n/2}}\big\{\Delta_n(x)+2^{-n}F_0(T)\big\},$$

where

$$\Delta_n(x):=K_n\big(La2^{n/2},x\big)-K_{n-1}\big(La2^{(n-1)/2},x\big).$$

Then we have for every $n\geq 0$ and $A\subseteq T$

$$e_n(A)\lesssim\left(a+\frac{1}{c}\right)\operatorname{diam}(A)+\sup_{x\in A}s_n(x).$$

PROOF. Let $n\geq 1$ and $b=e_n(A)/2>0$, else the result is trivial. Lemma 2.2 yields $N=2^{2^n}$ points $x_1,\dots,x_N\in A$ with $d(x_i,x_j)>b$ for $i\neq j$. Let

$$\sigma=\sup_{x\in A}s_n^a(x),\qquad r=\operatorname{diam}(A)+\sigma.$$

Case 1. If $\sigma>b/c$, then the conclusion is automatically satisfied as

$$e_n(A)<2c\sup_{x\in A}s_n^a(x)\lesssim\sup_{x\in A}s_n(x).$$

Case 2. If $\operatorname{diam}(A)>cb$, then the conclusion is automatically satisfied as

$$e_n(A)<\frac{2}{c}\operatorname{diam}(A).$$

Case 3. If $\sigma\leq b/c$ and $\operatorname{diam}(A)\leq cb$, then the sets $H_i=B(x_i,s_n^a(x_i))$, $i=1,\dots,N$ are (b,c)-separated, so the growth condition can be applied. We now repeat the proof of Lemma 6.4, except that we must pay the price $\Delta_n(x)$ for switching between two interpolation functionals [notice that $\Delta_n(x) \geq 0$ as F_n is a decreasing sequence of functionals]. To be precise, we estimate

$$\begin{aligned}
&F_0(T) - F_n(H_i) \\
&\quad \leq K_n\big(La2^{n/2}, x_i\big) + 2^{-n} F_0(T) \\
&\quad = K_{n-1}\big(La2^{(n-1)/2}, x_i\big) + \Delta_n(x_i) + 2^{-n} F_0(T) \\
&\quad \leq La2^{(n-1)/2} r + F_0(T) - F_{n-1}\big(B(x_i, r)\big) + \Delta_n(x_i) + 2^{-n} F_0(T) \\
&\quad \leq La2^{(n-1)/2} r + F_0(T) - F_{n-1}\left(\bigcup_{k \leq N} H_k\right) + \Delta_n(x_i) + 2^{-n} F_0(T)
\end{aligned}$$

for every $i \leq N$. Rearranging and applying the growth condition gives

$$\begin{aligned}
L2^{n/2} b &\leq F_{n-1}\left(\bigcup_{i \leq N} H_i\right) - \min_{i \leq N} F_n(H_i) \\
&\leq La2^{(n-1)/2} r + \sup_{x \in A} \Delta_n(x) + 2^{-n} F_0(T).
\end{aligned}$$

Using the definitions of b, r, Δ_n concludes the proof. $\square$

Note that the quantity $s_n(x)$ in Lemma 6.11 has an extra term as compared to Lemma 6.4. This is the price we pay for switching between different interpolation functionals. However, the additional term is innocuous: it gives rise to a telescoping sum when we apply the contraction principle.

PROOF OF THEOREM 6.9. Applying Lemma 6.11 and Theorem 3.1 yields

$$\gamma_2(T) \lesssim \left(a + \frac{1}{c}\right)\gamma_2(T) + \operatorname{diam}(T) + (a + c) \sup_{x \in T} \sum_{n \geq 1} 2^{n/2} s_n^a(x) + \frac{F_0(T)}{L},$$

where we used that $K_n(La2^{n/2}, x) \leq F_0(T)$ for every $n \geq 1$ and $x \in T$. Thus

$$\gamma_2(T) \lesssim \left(a + \frac{1}{c}\right)\gamma_2(T) + \frac{1 + c/a}{L} F_0(T) + \operatorname{diam}(T)$$

by Lemma 6.10. We can evidently choose a universal constant c_0 such that the conclusion of the theorem holds if $c \geq c_0$ and $a = 1/c_0$. $\square$

7. Dimension-free bounds on random matrices. As was stated in the Introduction, there are numerous challenging probabilistic problems that remain unsolved due to the lack of understanding of how to control the supremum of some concrete Gaussian process. Such problems arise, for example, in the study of structured random matrices [9, 13, 14], whose fine properties fall outside the reach of classical methods of random matrix theory. Problems of this kind constitute a particularly interesting case study for the control of inhomogeneous random processes, and provide concrete motivation for the development of new methods to control chaining functionals.

Of particular interest in the setting of structured random matrices are dimension-free bounds on matrix norms. Such bounds cannot be obtained by classical methods of random matrix theory such as the moment method, which are inherently dimension-dependent. This is explained in detail [13, 14] in the context of a tantalizing conjecture on Gaussian random matrices due to R. Latała. In this section, we make further progress in this direction by developing a closely related result: a dimension-free analogue of a well-known random matrix bound of M. Rudelson [7]. The proof provides another illustration of the utility of the contraction principle.

7.1. *Statement of results.* Let $A_1, \dots, A_m \in \mathbb{R}^{d\times d}$ be nonrandom symmetric matrices, and let $g_1, \dots, g_m$ be independent standard Gaussian variables. We are interested in bounding matrix norms of the random matrix

$$X = \sum_{k=1}^{m} g_k A_k$$

in terms of the coefficients A_k. A well-known result of M. Rudelson [7] (see also [11], Section 16.7, for a modern presentation) states that

$$\mathbf{E}\|X\| \lesssim \left\| \sum_{k=1}^{m} A_k^2 \right\|^{1/2} \sqrt{\log(m+1)}$$

in the important special case where each $A_k = x_k x_k^*$ has rank one (here and below $\|\cdot\|$ denotes the spectral norm of a matrix). Due to the rank-one assumption, the matrices A_k act nontrivially only on the m-dimensional subspace of $\mathbb{R}^d$ spanned by the vectors $x_1, \dots, x_m$, so that the above bound is overtly dimension-dependent. This dimension-dependence is not expected to be sharp when different vectors x_k possess substantially different scales. Unfortunately, the dependence on dimension arises in an apparently essential manner in the approach of [7]. We will see in the sequel that the contraction principle makes it possible to avoid this inefficiency. For example, we can obtain the following dimension-free form of Rudelson's bound.

THEOREM 7.1. *Suppose that each $A_k = x_k x_k^*$ has rank one. Then*

$$\mathbf{E}\|X\| \lesssim \left\| \sum_{k=1}^{m} A_k^2 \log(k+1) \right\|^{1/2}.$$

REMARK 7.2. Rudelson's bound was originally proved using a generic chaining construction. However, the generic chaining approach to Rudelson's dimension-dependent bound is essentially made obsolete by a much simpler and more general approach using the noncommutative Khintchine inequality of Lust-Piquard and Pisier [8]. The latter shows that an analogue of Rudelson's bound actually holds without any assumption on the matrices A_k (i.e., the rank-one assumption is not needed); see [14] for an elementary proof. However, it does not appear that such an approach could ever produce a dimension-free bound as in Theorem 7.1, as it relies crucially on the moment method of random matrix theory which is inherently dimension-dependent in nature [13]. In addition, the moment method is useless for bounding operator norms other than the spectral norm, which are important for applications in functional analysis [3, 4, 9]. Chaining methods appear to be essential for addressing problems of this kind that are out of reach of classical random matrix theory.

Theorem 7.1 arises as a special case of a much more general result that is of broader interest, and that clarifies the geometric structure behind the results of this section. In the remainder of this section, we will fix a symmetric compact convex set $B \subset \mathbb{R}^d$ that is 2-convex with constant η in the sense of Definition 5.7. We will be interested in controlling $\sup_{v\in T}\langle v, Xv\rangle$ for $T \subseteq B$. When $T = B = B_2^d$ is the Euclidean ball, this is simply the largest eigenvalue of X which is readily related to the spectral norm. However, we allow in general to consider any subset $T \subseteq B$. In addition, following [3, 4] we can consider any 2-convex ball B instead of the Euclidean ball, which will present no additional complications in the proofs.

As X is a Gaussian random matrix, clearly $v \mapsto \langle v, Xv\rangle$ is a centered Gaussian process. It therefore suffices by Theorem 1.1 to bound

$$\mathbf{E}\Big[\sup_{v\in T}\langle v, Xv\rangle\Big] \asymp \gamma_2(T, d),$$

where the natural distance $d(v, w)$ is given by

$$\begin{aligned} d(v, w) &:= \big[\mathbf{E}|\langle v, Xv\rangle - \langle w, Xw\rangle|^2\big]^{1/2} \\ &= \left[\sum_{k=1}^m \big\langle v + w, A_k(v - w)\big\rangle^2\right]^{1/2}. \end{aligned}$$

We will also define for $v, z \in \mathbb{R}^d$

$$\|v\|_z := \left[\sum_{k=1}^m \langle z, A_k v\rangle^2\right]^{1/2},$$

$$|||v||| := \left[\sum_{k=1}^m \langle v, A_k v\rangle^2\right]^{1/4}.$$

The main result of this section is the following, which could be viewed as a Gordon embedding theorem [11], Theorem 16.9.1, for structured matrices.

THEOREM 7.3. *Let $A_1, \dots, A_m$ be positive semidefinite. For any $T \subseteq B$,*

$$\mathbf{E}\Big[\sup_{v\in T}\langle v, Xv\rangle\Big] \lesssim \frac{1}{\sqrt{\eta}}\Big[\sup_{v\in T}\sum_{n\ge 0}\big(2^{n/2}e_n\big(B, \|\cdot\|_v\big)\big)^2\Big]^{1/2} + \gamma_{4,2}(T, |||\cdot|||)^2.$$

When Theorem 7.3 is specialized to the case $T = B = B_2^d$, we obtain a bound on the spectral norm of X from which Theorem 7.1 follows easily.

COROLLARY 7.4. *Let $A_1, \dots, A_m$ be positive semidefinite. Then*

$$\mathbf{E}\|X\| \lesssim \Big\|\sum_{k=1}^m A_k^2\Big\|^{1/2} + \sup_{n\ge 0} 2^{n/2}e_n\big(B_2^d, |||\cdot|||\big)^2.$$

The assumption that the matrices A_k are positive semidefinite is a natural relaxation of the rank-one assumption in Rudelson's approach [7]. This assumption ensures that $|||\cdot|||$ is a norm. Whether the positive semidefinite assumption can be weakened in Theorem 7.3 and Corollary 7.4 is a tantalizing question. Indeed, the above-mentioned conjecture of Latała [13] would follow if Corollary 7.4 were to hold for matrices A_k that are not positive semidefinite. While one can partially adapt the proof of Theorem 7.3 to general A_k, significant loss is incurred in the resulting bounds. These issues will be further discussed in Section 7.4 below.

7.2. *Proof of Theorem* 7.3. We will assume throughout this section that the matrices $A_1, \dots, A_m$ are positive semidefinite. This implies, in particular, that $|||\cdot|||$ is a norm and that $\|v\|_z \le |||v|||\,|||z|||$ by Cauchy–Schwarz.

Let us begin by explaining the basic geometric idea behind the proof through a back-of-the-envelope computation. Note that

$$d(y,z) = \|y-z\|_{y+z} \le 2\|y-z\|_x + |||y-z|||\big(|||y-x||| + |||z-x|||\big)$$

by the triangle inequality and Cauchy–Schwarz. Thus

$$\operatorname{diam}(A,d) \le 2\operatorname{diam}\big(A, \|\cdot\|_x\big) + 2\operatorname{diam}\big(A, |||\cdot|||\big)^2$$

for any $A \subseteq T$ and $x \in A$. This suggests we might try to bound $\gamma_2(T,d)$ by the sum of two terms, one of the form $\sup_{x\in T}\gamma_2(T, \|\cdot\|_x)$ and another of the form $\gamma_2(T, |||\cdot|||^2) = \gamma_{4,2}(T, |||\cdot|||)^2$. If that were possible, we would obtain a result far better than Theorem 7.3. The problem, however, lies with the first term: a direct application of the contraction principle would yield not $\sup_{x\in T}\gamma_2(T, \|\cdot\|_x)$, but rather the quantity

$$\inf_{(\mathcal{A}_n)}\sup_{x\in T}\sum_{n\ge 0} 2^{n/2}\operatorname{diam}\big(A_n(x), \|\cdot\|_x\big).$$

The latter could be much larger than $\sup_{x\in T}\gamma_2(T,\|\cdot\|_x)$: here, a single admissible sequence $(\mathcal{A}_n)$ must control simultaneously every norm $\|\cdot\|_x$, while in the definition of $\sup_{x\in T}\gamma_2(T,\|\cdot\|_x)$ each norm is controlled by its own admissible sequence. The remarkable aspect of Theorem 7.3 is that by exploiting the contraction theorem and 2-convexity of $B\supseteq T$, we will nonetheless achieve the same upper bound as would be obtained if we were to control $\sup_{x\in T}\gamma_2(B,\|\cdot\|_x)$ using Theorem 5.1.

We now proceed with the details of the proof. To exploit 2-convexity, it will be useful to replace the natural metric d by a regularized form:

$$\tilde{d}(v,w) := d(v,w) + |||v-w|||^2.$$

While $\tilde{d}$ is not a metric, it is a quasi-metric (the triangle inequality holds up to a multiplicative constant). This will suffice for all our purposes; in particular, it is readily verified that the proof of the contraction Theorem 3.1 holds verbatim in a quasi-metric space. We will use this observation in the sequel without further comment. The advantage of $\tilde{d}$, as opposed to the natural metric, is that it behaves in some sense like a norm.

LEMMA 7.5. *For every $v,w,z\in\mathbb{R}^d$, we have*:

(a) $\tilde{d}(v,w)\le 2(\tilde{d}(v,z)+\tilde{d}(z,w))$.
(b) $\tilde{d}(v,\frac{1}{2}(v+w))\le\frac{1}{2}\tilde{d}(v,w)$.

PROOF. The first claim follows from the triangle inequality and $(a+b)^2\le 2(a^2+b^2)$. To prove the second claim, note that we can write

$$v-\frac{1}{2}(v+w)=\frac{1}{2}(v-w),$$
$$v+\frac{1}{2}(v+w)=\frac{1}{2}(v-w)+(v+w).$$

Therefore,

$$\begin{aligned}\tilde{d}\biggl(v,\frac{1}{2}(v+w)\biggr)&=\frac{1}{2}\biggl\|\frac{1}{2}(v-w)+(v+w)\biggr\|_{v-w}+\frac{1}{4}|||v-w|||^2\\&\le\frac{1}{2}\bigl(\|v+w\|_{v-w}+|||v-w|||^2\bigr)\\&=\frac{1}{2}\tilde{d}(v,w),\end{aligned}$$

where we used the triangle inequality. $\square$

We now define the interpolation functional

$$K(t,x):=\inf_{y\in\mathbb{R}^d}\bigl\{\|y\|_B+t\tilde{d}(x,y)\bigr\},$$

and as usual we let $\pi_t(x)$ be a minimizer in this expression. Due to the second property of Lemma 7.5 (which was engineered precisely for this purpose), we can control the shrinkage of interpolation sets as in Theorem 5.8.

LEMMA 7.6. *Let $t \geq 0$ and $A \subseteq T$. Then $A_t := \{\pi_t(x) : x \in A\}$ satisfies*

$$A_t \subseteq z + \frac{L\sqrt{t}}{\sqrt{\eta}} \Big\{ \operatorname{diam}(A, \tilde{d}) + \sup_{x \in A} \tilde{d}\big(x, \pi_t(x)\big) \Big\}^{1/2} B$$

for some point in $z \in \mathbb{R}^d$, where L is a universal constant.

PROOF. Let $x \in A$ and $y = \pi_t(x)$. Then

$$\|y\|_B \leq K(t, x) \leq \|u\|_B + 2t\big(\tilde{d}(x, y) + \tilde{d}(y, u)\big)$$

for any $u \in \mathbb{R}^d$ by the definition of the interpolation functional and the first property of Lemma 7.5. Therefore, we have for every $y, z \in A_t$ and $u \in \mathbb{R}^d$

$$\max\big(\|y\|_B, \|z\|_B\big) \leq \|u\|_B + 2t \max\big(\tilde{d}(y, u), \tilde{d}(z, u)\big) + 2t \sup_{x \in A} \tilde{d}\big(x, \pi_t(x)\big).$$

If we choose $u = \frac{1}{2}(y + z)$, then we obtain

$$\max\big(\|y\|_B, \|z\|_B\big) \leq \left\| \frac{y + z}{2} \right\|_B + t\tilde{d}(y, z) + 2t \sup_{x \in A} \tilde{d}\big(x, \pi_t(x)\big)$$

using the second property of Lemma 7.5. In particular,

$$\eta \|y - z\|_B^2 \leq t\tilde{d}(y, z) + 2t \sup_{x \in A} \tilde{d}\big(x, \pi_t(x)\big)$$

for all $y, z \in A_t$ by 2-convexity of B. It follows that

$$\operatorname{diam}\big(A_t, \|\cdot\|_B\big) \leq \frac{\sqrt{t}}{\sqrt{\eta}} \Big\{ \operatorname{diam}(A_t, \tilde{d}) + 2 \sup_{x \in A} \tilde{d}\big(x, \pi_t(x)\big) \Big\}^{1/2}.$$

It remains to note that $\operatorname{diam}(A_t, \tilde{d}) \leq 4 \operatorname{diam}(A, \tilde{d}) + 8 \sup_{x \in A} \tilde{d}(x, \pi_t(x))$. $\square$

We now arrive at the main step in the proof of Theorem 7.3: we must verify the assumption of the contraction principle.

LEMMA 7.7. *Let $(\mathcal{C}_n)$ be an admissible sequence of T and $a, b > 0$. Then*

$$e_n(A, \tilde{d}) \lesssim b \operatorname{diam}(A, \tilde{d}) + \sup_{x \in A} s_n(x)$$

for every $n \geq 1$ and $A \subseteq T$, where

$$s_n(x) := (b + 1)\tilde{d}\big(x, \pi_{a2^{n/2}}(x)\big) + \frac{a 2^{n/2}}{b\eta} e_{n-1}\big(B, \|\cdot\|_x\big)^2 + \operatorname{diam}\big(C_{n-1}(x), |||\cdot|||\big)^2.$$

PROOF. Fix $n \geq 1$ and $A \subseteq T$. For every set $C \in \mathcal{C}_{n-1}$, define

$$A^C_{a2^{n/2}} := \{\pi_{a2^{n/2}}(x) : x \in A \cap C\}$$

and choose an arbitrary point $x_C \in A \cap C$. Now choose, for every $C \in \mathcal{C}_{n-1}$, a net $T^C_{n-1} \subseteq A^C_{a2^{n/2}}$ of cardinality less than $2^{2^{n-1}}$ such that

$$\inf_{z \in T^C_{n-1}} \|y - z\|_{x_C} \leq 4e_{n-1}(A^C_{a2^{n/2}}, \|\cdot\|_{x_C}) \qquad \text{for all } y \in A^C_{a2^{n/2}}.$$

Then $T_n := \bigcup_{C \in \mathcal{C}_{n-1}} T^C_{n-1}$ satisfies $|T_n| < 2^{2^n}$. It remains to show that

$$\sup_{x \in A} \tilde{d}(x, T_n) \lesssim b \operatorname{diam}(A, \tilde{d}) + \sup_{x \in A} s_n(x),$$

which concludes the proof.

To this end, fix $C \in \mathcal{C}_{n-1}$ and $x \in A \cap C$, and choose $z \in T^C_{n-1}$ such that

$$\|\pi_{a2^{n/2}}(x) - z\|_{x_C} \leq 4e_{n-1}(A^C_{a2^{n/2}}, \|\cdot\|_{x_C}).$$

We can estimate

$$\begin{aligned}
\tilde{d}(x, T_n) &\leq 2\tilde{d}(x, \pi_{a2^{n/2}}(x)) + 2\tilde{d}(\pi_{a2^{n/2}}(x), T_n) \\
&\leq 2\tilde{d}(x, \pi_{a2^{n/2}}(x)) + 2\tilde{d}(\pi_{a2^{n/2}}(x), z) \\
&\leq 2\tilde{d}(x, \pi_{a2^{n/2}}(x)) + 4\|\pi_{a2^{n/2}}(x) - z\|_{x_C} \\
&\quad + 2|||\pi_{a2^{n/2}}(x) - z|||(|||\pi_{a2^{n/2}}(x) - x_C||| + |||z - x_C|||) \\
&\quad + |||\pi_{a2^{n/2}}(x) - z|||^2.
\end{aligned}$$

As $z \in A^C_{a2^{n/2}}$ by construction, there is a point $x' \in A \cap C$ such that $z = \pi_{a2^{n/2}}(x')$. We therefore obtain, using that $|||v - w|||^2 \leq \tilde{d}(v, w)$,

$$\begin{aligned}
|||\pi_{a2^{n/2}}(x) - z||| &\leq |||x - \pi_{a2^{n/2}}(x)||| + |||x - x'||| + |||x' - \pi_{a2^{n/2}}(x')||| \\
&\leq 2 \sup_{v \in A} \tilde{d}(v, \pi_{a2^{n/2}}(v))^{1/2} + \operatorname{diam}(C, |||\cdot|||).
\end{aligned}$$

Similarly, we can estimate

$$|||\pi_{a2^{n/2}}(x) - x_C||| + |||z - x_C||| \leq 2 \sup_{v \in A} \tilde{d}(v, \pi_{a2^{n/2}}(v))^{1/2} + 2\operatorname{diam}(C, |||\cdot|||).$$

Putting together the above estimates, we obtain

$$\tilde{d}(x, T_n) \lesssim \sup_{v \in A} \tilde{d}(v, \pi_{a2^{n/2}}(v)) + e_{n-1}(A^C_{a2^{n/2}}, \|\cdot\|_{x_C}) + \operatorname{diam}(C, |||\cdot|||)^2$$

for every $x \in A \cap C$. We now note that by Lemma 7.6,

$$\begin{aligned}
&e_{n-1}(A^C_{a2^{n/2}}, \|\cdot\|_{x_C}) \\
&\quad\lesssim \frac{\sqrt{a2^{n/2}}}{\sqrt{\eta}}\Big\{\operatorname{diam}(A,\tilde{d}) + \sup_{v\in A}\tilde{d}\big(v,\pi_{a2^{n/2}}(v)\big)\Big\}^{1/2} e_{n-1}\big(B,\|\cdot\|_{x_C}\big) \\
&\quad\lesssim \frac{a2^{n/2}}{b\eta}\sup_{v\in A} e_{n-1}\big(B,\|\cdot\|_v\big)^2 + b\operatorname{diam}(A,\tilde{d}) + b\sup_{v\in A}\tilde{d}\big(v,\pi_{a2^{n/2}}(v)\big).
\end{aligned}$$

As $x \in A \cap C_{n-1}(x)$ for every $x \in A$, we have shown that

$$\begin{aligned}
\sup_{x\in A}\tilde{d}(x,T_n) &\lesssim b\operatorname{diam}(A,\tilde{d}) + (b+1)\sup_{v\in A}\tilde{d}\big(v,\pi_{a2^{n/2}}(v)\big) \\
&\qquad + \frac{a2^{n/2}}{b\eta}\sup_{v\in A} e_{n-1}\big(B,\|\cdot\|_v\big)^2 + \sup_{v\in A}\operatorname{diam}\big(C_{n-1}(v), \|\!|\cdot|\!\|\big)^2.
\end{aligned}$$

It remains to note that $\sup_v a_1(v) + \sup_v a_2(v) + \sup_v a_3(v) \le 3\sup_v(a_1(v) + a_2(v) + a_3(v))$ for any nonnegative functions $a_1(v), a_2(v), a_3(v) \ge 0$. $\square$

REMARK 7.8. We used above the standard fact that for any metric space (X,d) and $T \subseteq X$, there is a net $T_n \subseteq T$ with $|T_n| < 2^{2^n}$ so that $\sup_{x\in T} d(x,T_n) \le 4e_n(T,d)$. We recall the proof for completeness. The definition of entropy numbers guarantees the existence of a net $S_n \subseteq X$ with $|S_n| < 2^{2^n}$ so that $\sup_{x\in T} d(x,S_n) \le 2e_n(T,d)$, but S_n need not be a subset of T. For every point $z \in S_n$, choose $z' \in T$ such that $d(z,z') \le 2e_n(T,d)$, and let $T_n \subseteq T$ be the collection of points thus constructed. Then $d(x,T_n) \le d(x,S_n) + d(S_n,T_n) \le 4e_n(T,d)$ for every $x \in T$ as desired. The fact that one can choose the net T_n to be a subset of T rather than of X was essential in the above proof in order to ensure that $T^C_{n-1} \subseteq A^C_{a2^{n/2}}$.

We can now complete the proof of Theorem 7.3.

PROOF OF THEOREM 7.3. By Theorem 1.1, we have

$$\mathbf{E}\Big[\sup_{v\in T}\langle v, Xv\rangle\Big] \lesssim \gamma_2(T,d) \le \gamma_2(T,\tilde{d}).$$

Fix $a, b > 0$ and an admissible sequence $(\mathcal{C}_n)$ of T. Then

$$\begin{aligned}
\gamma_2(T,\tilde{d}) &\lesssim b\gamma_2(T,\tilde{d}) + \operatorname{diam}(T,\tilde{d}) + (b+1)\sup_{x\in T}\sum_{n\ge1} 2^{n/2}\tilde{d}\big(x,\pi_{a2^{n/2}}(x)\big) \\
&\qquad + \frac{a}{b\eta}\sup_{x\in T}\sum_{n\ge1}\big(2^{n/2}e_{n-1}\big(B,\|\cdot\|_x\big)\big)^2 \\
&\qquad + \sup_{x\in T}\sum_{n\ge1} 2^{n/2}\operatorname{diam}\big(C_{n-1}(x), \|\!|\cdot|\!\|\big)^2
\end{aligned}$$

by Theorem 3.1, where we used Lemma 7.7 to define $s_n(x)$ for $n \geq 1$ and the trivial choice $s_0(x) = \operatorname{diam}(T, \tilde{d})$. Choosing b to be a sufficiently small universal constant and applying the interpolation Lemma 4.5 gives

$$\gamma_2(T, \tilde{d}) \lesssim \operatorname{diam}(T, \tilde{d}) + \frac{1}{a} + \frac{a}{\eta} \sup_{x \in T} \sum_{n \geq 0} \big(2^{n/2} e_n(B, \|\cdot\|_x)\big)^2$$
$$+ \sup_{x \in T} \sum_{n \geq 0} 2^{n/2} \operatorname{diam}\big(C_n(x), |||\cdot|||\big)^2.$$

Optimizing over a and over admissible sequences $(\mathcal{C}_n)$ of T yields

$$\gamma_2(T, \tilde{d}) \lesssim \operatorname{diam}(T, \tilde{d}) + \frac{1}{\sqrt{\eta}} \left[\sup_{x \in T} \sum_{n \geq 0} \big(2^{n/2} e_n(B, \|\cdot\|_x)\big)^2 \right]^{1/2}$$
$$+ \gamma_{4,2}\big(T, |||\cdot|||\big)^2.$$

It remains to note that as $\operatorname{diam}(T, \tilde{d}) \leq 2 \operatorname{diam}(B, \|\cdot\|_x) + 2 \operatorname{diam}(T, |||\cdot|||)^2$ for any $x \in T$, the first term can be absorbed in the remaining two. $\square$

7.3. *Proofs of Corollary* 7.4 *and Theorem* 7.1. Using Theorem 7.3, the proof of Corollary 7.4 follows from classical entropy estimates for ellipsoids.

PROOF OF COROLLARY 7.4. Note that for any $v \in \mathbb{R}^d$, the norm $\|\cdot\|_v$ is a Euclidean norm defined by the inner product $\langle x, y \rangle_v := \langle x, \Sigma_v y \rangle$ with $\Sigma_v := \sum_{k=1}^m A_k v v^* A_k$. Thus $e_n(B_2^d, \|\cdot\|_v)$ are entropy numbers of ellipsoids in Hilbert space, which are well understood. Using the entropy estimates in [11], Section 2.5, we readily obtain

$$\sum_{n \geq 0} \big(2^{n/2} e_n(B_2^d, \|\cdot\|_v)\big)^2 \asymp \operatorname{Tr}[\Sigma_v] = \left\langle v, \left(\sum_{k=1}^m A_k^2 \right) v \right\rangle.$$

In particular, we obtain

$$\left[\sup_{v \in B_2^d} \sum_{n \geq 0} \big(2^{n/2} e_n(B_2^d, \|\cdot\|_v)\big)^2 \right]^{1/2} \asymp \left\| \sum_{k=1}^m A_k^2 \right\|^{1/2}.$$

On the other hand, by Theorem 5.8, we have

$$\gamma_{4,2}\big(B_2^d, |||\cdot|||\big) \asymp \sup_{n \geq 0} 2^{n/4} e_n\big(B_2^d, |||\cdot|||\big).$$

Thus Theorem 7.3 implies

$$\mathbf{E}\left[\sup_{v \in B_2^d} \langle v, Xv \rangle \right] \lesssim \left\| \sum_{k=1}^m A_k^2 \right\|^{1/2} + \sup_{n \geq 0} 2^{n/2} e_n\big(B_2^d, |||\cdot|||\big)^2.$$

It remains to note that

$$\|X\| = \sup_{v\in B_2^d} \big|\langle v, Xv\rangle\big| \le \sup_{v\in B_2^d} \langle v, Xv\rangle + \sup_{v\in B_2^d} \langle v, (-X)v\rangle$$

and that X and $-X$ have the same distribution. □

To deduce Theorem 7.1 from Corollary 7.4, we need to estimate the entropy numbers $e_n(B_2^d, \|\!|\cdot|\!\|)$. We will accomplish this using a classical result, the dual Sudakov inequality of N. Tomczak-Jaegermann ([11], Lemma 8.3.6).

PROOF OF THEOREM 7.1. We use the trivial estimate for $v \in B_2^d$

$$\|\!|v|\!\|^2 \le \|v\|_\sim := \sup_{z\in B_2^d} \|v\|_z.$$

This implies, using Remark 7.8 and the dual Sudakov inequality, that

$$e_n\big(B_2^d, \|\!|\cdot|\!\|\big)^2 \lesssim e_n\big(B_2^d, \|\cdot\|_\sim\big) \lesssim 2^{-n/2}\mathbf{E}\|g\|_\sim,$$

where g is a standard Gaussian vector in $\mathbb{R}^d$. Corollary 7.4 yields

$$\mathbf{E}\|X\| \lesssim \left\|\sum_{k=1}^m A_k^2\right\|^{1/2} + \mathbf{E}\|g\|_\sim.$$

Now suppose $A_k = x_k x_k^*$ have rank one. Then

$$\begin{aligned}
\mathbf{E}\|g\|_\sim &= \mathbf{E}\left[\sup_{z\in B_2^d} \sum_{k=1}^m \langle z, x_k\rangle^2 \langle x_k, g\rangle^2\right]^{1/2} \\
&\le \left[\sup_{z\in B_2^d} \sum_{k=1}^m \langle z, x_k\rangle^2 \|x_k\|^2 \log(k+1)\right]^{1/2} \mathbf{E}\left[\max_{k\le m} \frac{|\langle x_k, g\rangle|}{\|x_k\|\sqrt{\log(k+1)}}\right] \\
&\lesssim \left\|\sum_{k=1}^m A_k^2 \log(k+1)\right\|^{1/2},
\end{aligned}$$

using $A_k^2 = x_k x_k^* \|x_k\|^2$ and $\mathbf{E}[\max_k |G_k|/\sqrt{\log(k+1)}] \lesssim 1$ when G_k are (possibly dependent) standard Gaussians ([11], Proposition 2.4.16). □

7.4. *Discussion*. The aim of this section is to briefly discuss the connection between Corollary 7.4 and a conjecture of Latała. Let us briefly recall this conjecture, which is discussed in detail in [13]. Let X be a symmetric $d \times d$ matrix whose entries $\{X_{ij} : i \ge j\}$ are independent centered Gaussians with arbitrary variances

$X_{ij} \sim N(0, b_{ij}^2)$. Latała's conjecture states that the spectral norm of such a matrix is always of the same order as the maximum of the Euclidean norm of its rows,

$$\mathbf{E}\|X\| \overset{?}{\asymp} \mathbf{E}\left[\max_i \sqrt{\sum_j X_{ij}^2}\right].$$

The lower bound is trivial, as the spectral norm of any matrix is bounded below (deterministically) by the maximal Euclidean norm of its rows. It is far from obvious, however, why the upper bound should be true.

The independent entry model can be equivalently written as

$$X = \sum_{i \geq j} g_{ij} A_{ij}, \qquad A_{ij} = b_{ij}(e_i e_j^* + e_j e_i^*),$$

where $\{e_i\}$ denotes the standard basis in $\mathbb{R}^d$ and $\{g_{ij}\}$ are independent standard Gaussian variables. This model is therefore a special case of the general model considered in this section. Unfortunately, the matrices A_{ij} are not positive semidefinite. If the conclusion of Corollary 7.4 were to hold nonetheless for these matrices, then Latała's conjecture would follow readily. Indeed, arguing precisely as in the proof of Theorem 7.1, we would obtain

$$\begin{aligned}
\mathbf{E}\|X\| &\overset{?}{\lesssim} \left\|\sum_{i \geq j} A_{ij}^2\right\|^{1/2} + \mathbf{E}\left[\sup_{z \in B_2^d} \|g\|_z\right] \\
&\lesssim \max_i \sqrt{\sum_j b_{ij}^2} + \mathbf{E}\left[\max_i \sqrt{\sum_j b_{ij}^2 g_j^2}\right] \\
&\lesssim \mathbf{E}\left[\max_i \sqrt{\sum_j X_{ij}^2}\right],
\end{aligned}$$

where the last inequality was established in [13]. In view of these observations, it is of significant interest to understand to what extent the positive semidefinite assumption made in this section could be weakened.

An inspection of the proof of Theorem 7.3 shows that the positive semidefinite assumption was used only to ensure that $|||\cdot|||$ is a norm and that $\|v\|_z \leq |||v|||\,|||z|||$. All results in this section therefore continue to hold verbatim if we were to replace $|||\cdot|||$ in the statement and proof of Theorem 7.3 and Corollary 7.4 by an arbitrary (quasi)norm $|||\cdot|||'$ such that $\|v\|_z \lesssim |||v|||'\,|||z|||'$. This makes it possible, in principle, to prove much more general versions of these results. For example, the norm

$$|||v|||' = \left[\sum_{k=1}^m \langle v, |A_k| v\rangle^2\right]^{1/4}$$

satisfies the requisite condition for arbitrary $A_1, \ldots, A_m$, so that we obtain a general variant of Theorem 7.3 and Corollary 7.4 without any assumption on the coefficient matrices. However, significant loss is typically incurred when we naively replace A_k by $|A_k|$. For example, in the independent entry model this yields a bound of the form

$$\mathbf{E}\|X\| \lesssim \max_i \sqrt{\log i} \sqrt{\sum_j b_{ij}^2},$$

which is far larger than the bound suggested by Latała's conjecture.

Other choices of $|||\cdot|||'$ are possible in specific situations. For example, in the independent entry model, consider the choice

$$|||v|||' = \left[\sum_{i,j=1}^d v_i^2 b_{ij}^2 v_j^2\right]^{1/4}.$$

This defines a norm if we assume that the matrix of entry variances (b_{ij}^2) is positive semidefinite, in which case $\|v\|_z \lesssim |||v|||' |||z|||'$. This choice suffices to establish Latała's conjecture under the highly restrictive assumption that $(b_{ij}^2) \succeq 0$, recovering a result proved in [13] by different means.

In more general situations, it is not clear that it is possible to introduce a suitable (quasi)norm $|||\cdot|||'$ without incurring significant loss, and the resolution of Latała's conjecture will likely require additional geometric insight. Nonetheless, beside their independent interest, the results of this section provide a further step toward better understanding of the multiscale geometry of random matrices, and suggest that further development of the methods of this paper could yield new insights on various problems in this area.

REMARK 7.9. It is worth noting that even when $A_1, \ldots, A_m$ are positive definite, the geometric approach developed here is not necessarily efficient. Consider the trivial case where $m = 1$ and $A_1 = I$ is the identity matrix. Then obviously $\mathbf{E}\|X\| \asymp 1$, but Corollary 7.4 gives the terrible bound

$$\mathbf{E}\|X\| \lesssim 1 + \sup_{n\geq 0} 2^{n/2} e_n\big(B_2^d, \|\cdot\|_2\big)^2 \asymp \sqrt{d}.$$

Thus the geometric principle behind this section cannot fully explain the noncommutative Khintchine inequality discussed in Remark 7.2, though it actually improves on this inequality when the coefficient matrices have low rank. Discovering the correct geometric explanation of the noncommutative Khintchine inequality is closely related to another fundamental problem in the generic chaining theory, the convex hull problem ([11], pages 50–51), whose resolution may also shed new light on other random matrix problems (such as, e.g., the problem of obtaining sharp bounds in [9]).

Acknowledgments. The author is grateful to Richard Nickl for hosting a very pleasant visit to Cambridge during which some key results of this paper were obtained, to Roman Vershynin and Subhro Ghosh for motivating discussions, and to the referee and Associate Editor for helpful comments.

REFERENCES

[1] BALL, K., CARLEN, E. A. and LIEB, E. H. (1994). Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.* **115** 463–482. MR1262940

[2] CARL, B. (1981). Entropy numbers of diagonal operators with an application to eigenvalue problems. *J. Approx. Theory* **32** 135–150. MR0633698

[3] GUÉDON, O., MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2008). Majorizing measures and proportional subsets of bounded orthonormal systems. *Rev. Mat. Iberoam.* **24** 1075–1095. MR2490210

[4] GUÉDON, O. and RUDELSON, M. (2007). L_p-moments of random vectors via majorizing measures. *Adv. Math.* **208** 798–823. MR2304336

[5] LINDENSTRAUSS, J. and TZAFRIRI, L. (1979). *Classical Banach Spaces. II Function Spaces. Ergebnisse der Mathematik und Ihrer Grenzgebiete* [*Results in Mathematics and Related Areas*] **97**. Springer, Berlin. MR0540367

[6] NAOR, A. (2012). On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs. *Combin. Probab. Comput.* **21** 623–634. MR2942733

[7] RUDELSON, M. (1996). Random vectors in the isotropic position MSRI Preprint No. 1996-060.

[8] RUDELSON, M. (1999). Random vectors in the isotropic position. *J. Funct. Anal.* **164** 60–72. MR1694526

[9] RUDELSON, M. and VERSHYNIN, R. (2008). On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.* **61** 1025–1045. MR2417886

[10] TALAGRAND, M. (1996). Applying a theorem of Fernique. *Ann. Inst. Henri Poincaré Probab. Stat.* **32** 779–799. MR1422311

[11] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete* **60**. Springer, Heidelberg. MR3184689

[12] VAN HANDEL, R. (2016). Chaining, interpolation, and convexity. *J. Eur. Math. Soc.* (*JEMS*) To appear.

[13] VAN HANDEL, R. (2017). On the spectral norm of Gaussian random matrices. *Trans. Amer. Math. Soc.* **369** 8161–8178. MR3695857

[14] VAN HANDEL, R. (2017). Structured random matrices. In *Convexity and Concentration* (E. Carlen, M. Madiman and E. M. Werner, eds.). *IMA Vol. Math. Appl.* **161** 107–156. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
FINE HALL 207
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: rvan@princeton.edu