Discrete Comput Geom 34:365–379 (2005)
DOI: 10.1007/s00454-005-1186-y

Discrete & Computational
Geometry

The Geometry of Random $\{-1, 1\}$-Polytopes*

S. Mendelson,[1] A. Pajor,[1,2] and M. Rudelson[3]

[1]Centre for Mathematics and its Applications,
Institute of Advanced Studies, The Australian National University,
Canberra, ACT 0200, Australia
shahar.mendelson@anu.edu.au

[2]Laboratoire d'Analyse et Mathématiques Appliquées, Université de Marne-la-Vallée,
5 boulevard Descartes, Champs sur Marne, 77454 Marne-la-Vallée Cedex 2, France
alain.pajor@univ-mlv.fr

[3]Department of Mathematics, University of Missouri,
Columbia, MO 65211, USA
rudelson@math.missouri.edu

Communicated by Imre Bárány

Abstract. Random $\{-1, 1\}$-polytopes demonstrate extremal behavior with respect to many geometric characteristics. We illustrate this by showing that the combinatorial dimension, entropy and Gelfand numbers of these polytopes are extremal at every scale of their arguments.

1. Introduction

The goal of this article is to investigate some geometric properties of $\{-1, 1\}$-polytopes, which are symmetric convex hulls of subsets of the combinatorial cube $\{-1, 1\}^n$. Formally, let $n \geq 1$ and $N \geq 1$ be integers. For any set $\{\omega_i \colon 1 \leq i \leq N\} \subset \{-1, 1\}^n$, define

$$K_{n,N} = K_{n,N}(\omega_1, \ldots, \omega_N) = \mathrm{conv}(\pm\omega_1, \ldots, \pm\omega_N) = \mathrm{absconv}(\omega_1, \ldots, \omega_N).$$

* The research by the first author was supported in part by an Australian Research Council Discovery grant, that by the second author was supported in part by an Australian Research Council Fellowship, and the third author's research was supported in part by NSF Grant DMS-024380.

Our focus is on *random* $\{-1, 1\}$-polytopes, where the randomness is generated by the uniform (counting) probability measure on $\{-1, 1\}^n$. We say that a certain property is satisfied by a random $\{-1, 1\}$-polytope if the set of polytopes $K_{n,N}$ satisfying this property has probability larger than $1 - c^n$, where $c \in (0, 1)$ is a numerical constant which is independent of n and N.

Equivalently, one can consider the random structure at hand in the following manner. Let ξ be a symmetric $\{-1, 1\}$-valued random variable and let $(\xi_{i,j})$, $1 \le i \le N$, $1 \le j \le n$, be independent copies of ξ. If $e_1, \dots, e_n$ denote the standard unit vectors, each $X_i = \sum_{j=1}^n \xi_{i,j} e_j$ is a random point in $\{-1, 1\}^n$ and $K_{n,N} = \text{absconv}(X_1, \dots, X_N)$.

Throughout this article, we denote by $\| \cdot \|$ the canonical Euclidean norm. The corresponding unit ball and its unit sphere are denoted by B_2^n and S^{n-1}, respectively. For any Lebesgue measurable set $L \subset \mathbb{R}^n$, put $\text{vol}(L)$ to be the volume of L and for a set $T \subset \mathbb{R}^n$, let $\text{absconv}(T)$ be its symmetric convex hull.

It is well known that random polytopes generated by random points on the sphere demonstrate the extremal behavior with respect to many geometric characteristics (see for instance [23] and an extensive survey [12]). The investigation of the complexity of random $\{-1, 1\}$-polytopes or, equivalently, 0/1-polytopes is more recent (see the survey [26]). For example, see [3] for the study of the number of facets and [9] where it is established that the volume of a random $\{-1, 1\}$-polytope with N vertices is the largest possible among all polytopes $K_{n,N}$.

The main results of this article show that this extremal behavior is true for three important geometric parameters—the combinatorial dimension, the entropy and the Gelfand numbers (defined below). All three parameters are scale-sensitive, and our results show that random polytopes are the "worst possible" among all polytopes $K_{n,N}$ at every scale of the parameter in question. Indeed, we show that the behavior in the random case matches the upper bounds that hold for any polytope $K_{n,N}$.

The significance of such results is the fact that the parameters in question play a central role in Asymptotic Geometric Analysis, Empirical Processes theory and Non-parametric Statistics (see, e.g., [1], [8], [14], [16]–[18] and references therein), where they serve as a way of measuring the richness or the complexity of a given set. Hence, our result is yet another indication that random polytopes are the "most complicated" in the class $K_{n,N}$.

Definition 1.1. Let (Y, d) be a metric space and let $K \subset Y$. For every $\varepsilon > 0$, we define the covering number $N(K, \varepsilon, d)$ at scale ε, as the minimal number of balls of radius ε (with respect to the metric d) needed to cover K.

Usually, we use the Euclidean metric in $\mathbb{R}^n$, in which case, for any $\varepsilon > 0$, we denote the covering number at scale ε by $N(K, \varepsilon B_2^n)$, that is, the number of translates of the n-dimensional Euclidean ball of radius ε needed to cover K. More generally, $N(A, B)$ is the number of translates of B needed to cover A.

Definition 1.2. Let (Y, d) be a metric space. A set is ε-separated with respect to a metric d if the distance between every two distinct points in the set is larger than ε. We denote the maximal cardinality of an ε-separated subset of Y by $D(Y, \varepsilon, d)$.

As for the covering numbers, when using the Euclidean metric on a set K of $\mathbb{R}^n$, we denote by $D(K, \varepsilon B_2^n)$ the maximal cardinality of an ε-separated subset of K.

It is easy to see that the cardinality of a maximal ε-separated subset of Y is equivalent to the covering numbers of Y, namely, for every $\varepsilon > 0$, $N(Y, \varepsilon, d) \le D(Y, \varepsilon, d) \le N(Y, \varepsilon/2, d)$.

The second parameter we study is the combinatorial dimension, which measures the tradeoff between the size of a cube contained in a coordinate projection of a set F and the dimension of the projection.

This parameter was introduced independently by several authors—particularly in the context of empirical processes (see, for example, [17] and [22]).

Definition 1.3. Let F be a set of functions $f\colon \Omega \to \mathbb{R}$. For every $\varepsilon > 0$, a set $\sigma = \{x_1, \dots, x_n\} \subset \Omega$ is said to be ε-shattered by F if there is some function $s\colon \sigma \to \mathbb{R}$, such that for every $I \subset \{1, \dots, n\}$ there is some $f_I \in F$ for which $f_I(x_i) \ge s(x_i) + \varepsilon$ if $i \in I$, and $f_I(x_i) \le s(x_i) - \varepsilon$ if $i \notin I$. Define the shattering dimension at scale ε as

$$\mathrm{VC}(F, \Omega, \varepsilon) = \sup\{|\sigma| \mid \sigma \subset \Omega,\ \sigma \text{ is } \varepsilon\text{-shattered by } F\},$$

where $|\sigma|$ denotes the cardinality of σ.

In cases where the underlying space is clear we denote the combinatorial dimension by $\mathrm{VC}(F, \varepsilon)$. If F is $\{-1, 1\}$-valued, we denote its combinatorial dimension by $\mathrm{VC}(F)$.

Observe that the combinatorial dimension is a scale-sensitive version of the Vapnik–Chervonenkis (VC) dimension [25], which is defined for subsets of the combinatorial cube as the largest dimension of a coordinate projection of F which is the entire combinatorial cube of that dimension.

In our case the underlaying space will always be the set of coordinates given by the standard unit basis $\{e_1, \dots, e_n\}$ and each vector in $\mathbb{R}^n$ is a function on this set in the natural way. Also, since we are only interested in convex symmetric sets (as $F = K_{n,N}$ is convex and symmetric), it is possible to take the level function $s \equiv 0$ (see, e.g. [13]). Hence, the combinatorial dimension of $K_{n,N}$ at scale ε is simply the largest dimension of a subset $\sigma \subset \{1, \dots, n\}$ such that the coordinate projection P_σ from $\mathbb{R}^n$ onto $\mathbb{R}^\sigma$ satisfies

$$\varepsilon B_\infty^{|\sigma|} \subset P_\sigma K_{n,N} = \{(k(i))_{i \in \sigma}\colon k \in K_{n,N}\},$$

where B_∞^d is the cube of dimension d.

Since our results only hold for a certain range of N and n, we require the following assumption:

Assumption 1. $2n \le N \le 2^n$.

A result we use throughout this article was recently proved in [11], and shows that a random polytope contains the interpolation body generated by the cube and a "large" Euclidean ball.

Theorem 1.4. *There exist absolute positive constants c, c_1 and c_2 for which the following holds. Let n and N be integers such that $n < N \le 2^n$ and let $\alpha = \alpha(N,n) = n/(N-n)$. For every $0 < \beta \le \frac{1}{2}$ one has*

$$\Pr\left(\left\{K_{n,N} \supset C(\alpha)\left(\sqrt{\beta \log(2N/n)}\, B_2^n \cap B_\infty^n\right)\right\}\right) \ge 1 - \exp(-cn^\beta N^{1-\beta}),$$

where $C(\alpha) = c_1 c_2^\alpha$.

We mention that a similar result was obtained by Giannopoulos and Hartzoulaki [9], though for a slightly more restrictive range of N, namely, for $N \ge n \log 2n$, and with a weaker probability estimate—only $1 - \exp(-cn)$.

Observe that

$$C(\alpha)\left(\sqrt{\beta \log(2N/n)} B_2^n \cap B_\infty^n\right) \supset C(\alpha)\sqrt{\frac{\beta \log(2N/n)}{n}}\, B_\infty^n,$$

and, in particular, Theorem 1.4 implies that if Assumption 1 is satisfied and indeed $N \ge 2n$, then with probability at least $1 - \exp(-cn^\beta N^{1-\beta})$,

$$\mathrm{vol}^{1/n}(K_{n,N}) \ge c_1 \sqrt{\frac{\beta \log(2N/n)}{n}} \tag{1}$$

for some absolute constants c and c_1.

The article is organized as follows. The next section is devoted to the proof of some deterministic upper bounds on the entropy and the combinatorial dimension of symmetric convex hulls of subsets of cardinality N of $\sqrt{n} S^{n-1}$; hence, these estimates hold true for any $\{-1, 1\}$-polytope. In particular, we prove a complementary result to the Carl–Pajor theorem [6], by obtaining an entropy estimate for scales smaller than $c\sqrt{\log(N/n)}$. In Section 3 we show that both upper bounds are sharp as they are attained by a random $\{-1, 1\}$-polytope in both cases. We end the article by proving a similar result for Gelfand numbers (defined below).

Finally, a notational convention. Throughout, all absolute constants are positive numbers and are denoted by c, C, K and κ. Their values may change from line to line, or even in the same line. We write $a \sim b$ if there are absolute positive constants c and C such that $ca \le b \le Ca$.

2. Deterministic Upper Bounds

The first deterministic upper bound we require is on the ℓ_2^n entropy of any $\{-1, 1\}$-polytope, and was established in [6].

Theorem 2.1. *There exist absolute positive constants c_0 and c_1 for which the following holds. Let $N \ge n$, let $T \subset \sqrt{n} S^{n-1}$ with $|T| \le N$ and put $K = \mathrm{absconv}(T)$. Then, for any $\varepsilon \ge c_0\sqrt{n/N}$,*

$$\log N(K, \varepsilon B_2^n) \le c_1 \frac{n}{\varepsilon^2} \log\left(\frac{c_1 N \varepsilon^2}{n}\right).$$

A result of a similar flavor is a volumetric estimate on K, which was established independently in [2], [6] and [10].

Theorem 2.2. *There exists an absolute positive constant c such that for any K as above*,

$$\mathrm{vol}(K)^{1/n} \le c\left(\frac{\log(cN/n)}{n}\right)^{1/2}.$$

An immediate corollary which follows from Theorem 2.2 is an estimate on the combinatorial dimension of any $\{-1,1\}$-polytope.

Corollary 2.3. *There exists an absolute positive constant C such that for any polytope $K_{n,N}$ and any $0 < \varepsilon \le 1$*,

$$\mathrm{VC}(K_{n,N},\varepsilon) \le \min\left\{C\,\frac{\log(CN\varepsilon^2)}{\varepsilon^2}, n\right\}.$$

Proof. Since a projection onto k coordinates of a $\{-1,1\}$-polytope in $\mathbb{R}^n$ is a $\{-1,1\}$-polytope in $\mathbb{R}^k$, then by the volumetric estimate of Theorem 2.2, it is clear that a k-projection of such a polytope cannot contain rB_∞^k for r larger than $c(\log(N/k)/k)^{1/2}$, from which the estimate easily follows. □

It is evident from the formulation of Theorem 2.1 that it is not optimal for all scales of ε. The main result of this section is an entropy estimate for any polytope $K_{n,N}$ and $\varepsilon \le c\sqrt{\log(N/n)}$. This estimate will later be shown to be sharp.

Theorem 2.4. *There exist absolute positive constants c_0 and c_1 for which the following holds. Let $T \subset \sqrt{n}S^{n-1}$ with $|T| \le N$ and set K to be its symmetric convex hull. Then for any $\varepsilon \le \sqrt{\log(c_0N/n)}$*,

$$\log N(K,\varepsilon B_2^n) \le n\log\left(\frac{c_1\sqrt{\log(c_1N/n)}}{\varepsilon}\right).$$

Before presenting the proof, we introduce some volumetric parameters of a convex body K which are related to its mixed volumes (see [18] and [20]).

Definition 2.5. Let K be a convex compact subset of $\mathbb{R}^n$. For every $1 \le d \le n$, set

$$w_d(K) = \left(\frac{1}{\mathrm{vol}(B_2^d)}\int_{G_{n,d}} \mathrm{vol}(P_EK)\,dE\right)^{1/d},$$

where P_E is the orthogonal projection onto E and dE is the Haar probability measure on the Grassmann manifold of subspaces of dimension d of $\mathbb{R}^n$. We also set $w_0(K) = 1$.

The well-known Alexandrov inequalities state that for $1 \le d \le n$, $w_d(K)$ is non-increasing.

For a symmetric convex body K, let $K^\circ = \{x\colon \langle x, y\rangle \le 1 \text{ for any } y \in K\}$, where $\langle x, y\rangle$ denotes the scalar product of vectors x and y. Set $M^*(K) = \int_{S^{n-1}} \|x\|_{K^*}\, d\sigma$, where σ is the Haar probability measure on the sphere and $\|\cdot\|_{K^*}$ is the norm for which K° is its unit ball. It is easy to verify that $w_1(K) = M^*(K)$, and thus, for $1 \le d \le n$, $w_d(K) \le M^*(K)$ (see Chapter 9 of [18] or Chapter 6 of [20]).

Finally, recall the Steiner–Minkowski formula (see [18] and [20]), that for any $t > 0$,

$$\frac{\mathrm{vol}(K + tB_2^n)}{\mathrm{vol}(B_2^n)} = \sum_{d=0}^{n} \binom{n}{d} t^{n-d} w_d^d(K). \tag{2}$$

Lemma 2.6. *Let T and K be as in Theorem* 2.4. *Then for every* $1 \le d \le n$,

$$w_d(K) \le c\sqrt{\log\left(\frac{cN}{d}\right)},$$

where c is an absolute positive constant.

Proof. Fix $1 \le d \le n$ and for $u \ge 1$ set

$$\Omega_u = \left\{E \in G_{n,d}\colon u\sqrt{d} \le \sup_{t\in T} \|P_E t\| < (u+1)\sqrt{d}\right\}.$$

By a standard concentration argument for Lipschitz functions on the sphere and the connection between the Haar measure on the sphere and on the Grassmann manifold [16], there exists $\kappa > 0$, such that for every $d \ge \kappa \log N$ and $u \ge 1$, $Pr(\Omega_{u+1}) \le \exp(-c_0 u^2 d)$. Applying Theorem 2.2, it is evident that if $T \subset \sqrt{d} B_2^d$ and $|T| \le N$ then $\mathrm{vol}\,(\mathrm{absconv}(T)) \le c^d (\log(cN/d)/d)^{d/2}$. Hence, if $E \in \Omega_u$ then

$$P_E K = \mathrm{absconv}(P_E T) \subset (u+1)\sqrt{d} B_2^d,$$

and

$$\begin{aligned}\int_{G_{n,d}} \mathrm{vol}(P_E K)\, dE \;&\le\; \int_{\Omega_0} \mathrm{vol}(P_E K)\, dE + \sum_{u=1}^{\infty} \int_{\Omega_u} \mathrm{vol}(P_E K)\, dE \\ &\le\; c^d \left(\frac{\log(cN/d)}{d}\right)^{d/2} \left(1 + \sum_{u=1}^{\infty} (u+1)^d \exp(-cu^2 d)\right) \\ &\le\; c_1^d \left(\frac{\log(c_1 N/d)}{d}\right)^{d/2}.\end{aligned}$$

The claim now follows for $d \ge \kappa \log N$ because $\mathrm{vol}(B_2^d)^{1/d} \sim 1/\sqrt{d}$.

It is well known (see, for instance, Lemma 4.14 of [18]) that if $T \subset \sqrt{n} S^{n-1}$ and $|T| \le N$ then $M^*(K) \le c_2\sqrt{\log N}$, and since $w_d(K) \le M^*(K)$ then for $d \le \kappa \log N$,

$$w_d(K) \le M^*(K) \le c_2\sqrt{\log N} \le c_3\sqrt{\log\left(\frac{c_3 N}{d}\right)},$$

which concludes the proof. □

Proof of Theorem 2.4. It is standard to verify that if A and B are convex and symmetric sets in $\mathbb{R}^n$ and $B \subset A$ then $N(A, B) \le 3^n \operatorname{vol}(A)/\operatorname{vol}(B)$ (see [18]). In particular,

$$N(K, \varepsilon B_2^n) \le N(K + \varepsilon B_2^n, \varepsilon B_2^n) \le 3^n \frac{\operatorname{vol}(K + \varepsilon B_2^n)}{\operatorname{vol}(\varepsilon B_2^n)}.$$

By the Steiner–Minkowski formula (2) and the previous lemma,

$$\begin{aligned}\frac{\operatorname{vol}((1/\varepsilon)K + B_2^n)}{\operatorname{vol}(B_2^n)} &= \sum_{d=0}^{n} \binom{n}{d} \left(\frac{w_d(K)}{\varepsilon}\right)^d \le \sum_{d=0}^{n} \binom{n}{d} \left(\frac{c^2}{\varepsilon^2} \log\left(\frac{cN}{d}\right)\right)^{d/2} \\ &= \sum_{d=0}^{n} \binom{n}{d} \rho_d,\end{aligned}$$

where $\rho_d = ((c^2/\varepsilon^2)\log(cN/d))^{d/2}$. A straightforward computation shows that there exists an absolute positive constant c_3 such that if $\varepsilon \le \sqrt{\log(cN/n)}$, then for every $1 \le d \le n$ and every N and ε,

$$\rho_d \le \left(c_3 \frac{\log(c_3 N/n)}{\varepsilon^2}\right)^{n/2}.$$

Hence, for some absolute constant c_4, we have

$$\log N(K, \varepsilon B_2^n) \le n \log\left(c_4 \frac{\sqrt{\log(c_4 N/n)}}{\varepsilon}\right),$$

as claimed. □

It is convenient to use the terminology of the so-called s-numbers (see [18]). For a subset $K \subset \mathbb{R}^n$ and any $j \ge 1$, the jth Gelfand number is defined by $c_j(K) = \inf\{\max_{x \in K \cap E} \|x\| \colon E \text{ subspace of } \mathbb{R}^n, \operatorname{codim}(E) < j\}$ and the jth entropy number is defined by $e_j(K) = \inf\{\varepsilon\colon N(K, \varepsilon B_2^n) \le 2^{j-1}\}$. Thus, the kth Gelfand number of a body is half of the smallest diameter of a $(k-1)$-codimensional section of K and the entropy numbers are the discrete inverse of the logarithm of the covering numbers.

Just like the upper bound on the entropy (and thus on e_k), one can prove the following upper estimate on the Gelfand numbers.

Theorem 2.7 [6]. *There exists an absolute positive constants c_0 such that the following holds. Let $N \ge n$, let $T \subset \sqrt{n}\, B_2^n$ and put $K = \operatorname{absconv}(T)$. Then, for any $1 \le k \le n$,*

$$c_k(K) \le c_0 \min\left\{\sqrt{n}, \left(\frac{n \log(2N/k)}{k}\right)^{1/2}\right\}.$$

3. Lower Bounds for Random Polytopes

We start by formulating and proving the lower bound on the combinatorial dimension of a random polytope.

Theorem 3.1. *There exist absolute positive constants c and c_1 for which the following holds. Let n and N be integers which satisfy Assumption* 1. *Then, for any* $0 < \beta < \frac{1}{2}$ *and* $N \geq n$, *with probability of at least* $1 - \exp(-cn^\beta N^{1-\beta})$, *for every* $0 < \varepsilon < 1$,

$$\mathrm{VC}(K_{n,N}, \varepsilon) \geq \min\left\{C_\beta \frac{\log(c_1 N \varepsilon^2)}{\varepsilon^2}, n\right\},$$

where C_β depends only on β.

A well-known bound on the cardinality of subsets of the combinatorial cube is the Sauer–Shelah lemma [19], [21], [25].

Theorem 3.2. *If* $T \subset \{-1, 1\}^n$ *and* $d = \mathrm{VC}(T)$, *then*

$$|T| \leq \sum_{i=0}^{d} \binom{n}{i} \leq \left(\frac{en}{d}\right)^d,$$

where the last inequality holds if $n \geq d$. *In particular, if* $|T| \geq 2^{\alpha n}$ *then* $\mathrm{VC}(T) \geq C_\alpha n$, *where C_α depends only on α.*

Proof of Theorem 3.1. We prove a lower bound on the inverse function of the combinatorial dimension of a convex symmetric set A. For $1 \leq d \leq n$, let $f_A(d)$ be the largest ε such that, for some $\sigma \subset \{1, \ldots, n\}$ with $|\sigma| = d$, $\varepsilon B_\infty^d \subset P_\sigma A = \{(a(i))_{i\in\sigma} : a \in A\}$. Clearly, our claim will follow if we show that with high probability, for any $1 \leq d \leq n$, $f_{K_{n,N}}(d) \geq \min\{C_\beta \sqrt{\log(2N/d)/d}, 1\}$.

First, suppose that $4d \leq \log_2 N$ and divide the set $\{1, \ldots, n\}$ into subsets of cardinality $2d$. Consider one of these subsets, say $J = \{1, \ldots, 2d\}$, and denote by P_J the coordinate projection from $\mathbb{R}^n$ onto $\mathbb{R}^J$. Let $T_{n,N}$ be the set of vertices of $K_{n,N}$. Then

$$Pr(\{|P_J T_{n,N}| \leq 2^{2d-1}\}) \leq \sum_{\ell=1}^{2^{2d-1}} \binom{2^{2d}}{\ell} \cdot \left(\frac{\ell}{2^{2d}}\right)^N \leq 2^{2^{2d}} \cdot \left(\frac{1}{2}\right)^N.$$

Since $4d \leq \log_2 N$, the last expression does not exceed $2^{-N/2}$. Note that the projections $P_J T_{n,N}$ are independent for disjoint subsets J, so the probability that all such projections contain less than 2^{2d-1} distinct elements is at most $2^{-(n/2d)N/2}$. Assume now that the projection on at least one subset J contains more than 2^{2d-1} elements. By the Sauer–Shelah lemma, $\mathrm{VC}(P_J T_{n,N}) \geq d$ and thus $\mathrm{VC}(T_{n,N}) \geq d$. Therefore, when $4d \leq \log_2 N$, we have $f_{K_{n,N}}(d) \geq 1$ with probability higher than $1 - 2^{-(n/2d)N/2} \geq 1 - \exp(-cn^\beta N^{1-\beta})$ for some absolute constant c.

Next, fix $d \geq \log_2 N$ and thus $2^d \geq N \geq 2n \geq 2d$. Again, we divide $\{1, \ldots, n\}$ into disjoint subsets with d elements, and since the coordinate projections onto these subsets are "independent" random $K_{d,N}$ polytopes, then by Theorem 1.4 at least one of these polytopes contains a cube of size $C\sqrt{\beta \log(2N/d)/d}$ with probability greater than

$$1 - \exp(-c(n/d)d^\beta N^{1-\beta}) \geq 1 - \exp(-cn^\beta N^{1-\beta}).$$

Hence, with that probability, $f_{K_{n,N}}(d) \geq C\sqrt{\beta \log(2N/d)/d}$.

Since the function $f_{K_{n,N}}$ is non-increasing, for $\frac{1}{4}\log_2 N \le d < \log_2 N$ and $1 \le d \le n$, we have

$$f_{K_{n,N}}(d) \ge f_{K_{n,N}}(\log_2 N) \ge c \ge C\sqrt{\frac{\beta \log(2N/d)}{d}}$$

with probability at least $1 - \exp(-cn^{\beta}N^{1-\beta})$. □

Theorem 3.1 can be used to resolve the following question. It was shown in [15] that there are absolute positive constants c and C such that for any class of functions bounded by 1,

$$\mathrm{VC}(\mathrm{conv}(F), \varepsilon) \le C \cdot \frac{\mathrm{VC}(F, c\varepsilon)}{\varepsilon^2}.$$

It was also shown that this estimate is sharp up to a logarithmic factor, in the following sense:

Theorem 3.3. *There exist absolute positive constants C and c for which the following holds. For every $0 < \varepsilon < \frac{1}{2}$ there is a class F_ε of functions bounded by 1 such that*

$$\mathrm{VC}(\mathrm{conv}(F_\varepsilon), \varepsilon) \ge C \cdot \frac{\mathrm{VC}(F_\varepsilon, c\varepsilon)}{\varepsilon^2 \log(1/\varepsilon)}.$$

Now, one can remove the logarithmic factor and construct a set for which the lower bound matches the upper one for "most" values of ε.

Theorem 3.4. *There exist absolute positive constants c_1 and c_2 for which the following holds. Let T be a random subset of $\{-1, 1\}^n$ with $2n$ elements and set $F = T \cup -T$. Then, with probability at least $1 - \exp(-c_1 n)$, for any $\gamma < \frac{1}{2}$ and $\varepsilon \ge c_2/n^{\gamma}$,*

$$\mathrm{VC}(\mathrm{conv}(F), \varepsilon) \ge c_3(\gamma) \cdot \frac{\mathrm{VC}(F, \varepsilon)}{\varepsilon^2},$$

where $c_3(\gamma)$ depends only on γ.

Proof. Since F consists of $\{-1, 1\}$-valued functions (on the coordinates $\{e_1, \dots, e_n\}$), then for any $\varepsilon > 0$, $\mathrm{VC}(F, \varepsilon) \le c \log n$. On the other hand, by Theorem 3.1 for $\beta = \frac{1}{2}$, with probability at least $1 - \exp(-cn^{\beta}N^{1-\beta}) \ge 1 - \exp(-cn)$ for any $\gamma < \frac{1}{2}$ and $\varepsilon \ge c_2/n^{\gamma}$,

$$\mathrm{VC}(\mathrm{conv}(F), \varepsilon) \ge \frac{c}{\varepsilon^2}\log(cn\varepsilon^2) \ge \frac{c'(\gamma)}{\varepsilon^2}\log n \ge c_3(\gamma)\frac{\mathrm{VC}(F, \varepsilon)}{\varepsilon^2}. \qquad \square$$

Next, we turn to the question of entropy. We will show that at a scale below $c\sqrt{\log(N/n)}$, a lower bound on the entropy follows from the fact that $K_{n,N}$ contains the interpolation body $\alpha B_2^n \cap B_\infty^n$ for an appropriate value of α, and thus must have a large entropy. However, for larger scales, one needs an additional argument in order to construct a large separated subset in $K_{n,N}$.

Theorem 3.5. *There exist absolute positive constants C, κ, c, c_1 and c_2 for which the following holds. For any $\kappa\sqrt{\log(N/n)} \le \varepsilon \le C\sqrt{n}$, with probability at least $1 - \exp(-cn)$,*

$$\log D(K_{n,N}, \varepsilon B_2^n) \ge c_1 \frac{n}{\varepsilon^2} \log\left(\frac{c_2 N \varepsilon^2}{n}\right).$$

The proof of the theorem requires some preparation.

Lemma 3.6. *Let $0 < \lambda \le \frac{1}{2}$ and for every integer N fix $m \le N/2$. Let $B(N,m)$ be the family of subsets of $\{1, \dots, N\}$ of cardinality m. Then there exists a subset $P \subset B(N,m)$ which satisfies that $\log|P| \ge (1-\lambda)m \log(c_\lambda(N/n))$ and if $I, J \in P$ and $I \ne J$ then $|I \bigtriangleup J| \ge \lambda m$. In other words,*

$$\log D(B(N,m), \lambda m, d_{\mathrm{H}}) \ge (1-\lambda)m \log\left(c_\lambda \frac{N}{m}\right),$$

where d_{H} is the Hamming metric (that is, $d_{\mathrm{H}}(I,J) = |I \bigtriangleup J|$).

Proof. Without loss of generality, assume that λm is an integer. Pick any subset of cardinality m of $\{1, \dots, N\}$ and throw away all subsets of size m such that $|I \bigtriangleup J| \le \lambda m$. There are at most

$$\sum_{k=(1-\lambda)m}^{m} \binom{m}{k}\binom{N-m}{m-k} \le 2^m \max_{(1-\lambda)m \le k \le m} \binom{N-m}{m-k} \le 2^m \binom{N}{\lambda m}$$

such subsets, since $m \le N/2$. Now, select a new subset of size m from the remaining subsets. Repeating this argument, we obtain a family P of subsets of size m which are λm-separated in the Hamming metric and with cardinality larger than

$$\binom{N}{m} \Big/ 2^m \binom{N}{\lambda m} \ge \frac{(N/2m)^m}{2^m (Ne/\lambda m)^{\lambda m}},$$

which concludes the proof. □

Next, we use the following formulation of Bernstein's inequality:

Theorem 3.7 [4], [24]. *Let $Z_1, \dots, Z_n$ be independent random variables with zero mean, such that for every i and every $k \ge 2$, $\mathbb{E}|Z_i|^k \le k!\, M^{k-2} v_i/2$. Then, for any $v \ge \sum_{i=1}^n v_i$ and any $u > 0$,*

$$Pr\left(\left\{\left|\sum_{i=i}^n Z_i\right| > u\right\}\right) \le 2\exp\left(-\frac{u2}{2(v+uM)}\right).$$

One can formulate Theorem 3.7 using the ψ_1 norm of the random variable Z. Recall that $\|Z\|_{\psi_1} = \inf\{b > 0\colon \mathbb{E}\exp(|Z|/b) \le 2\}$. Random variables with a bounded ψ_1 norm display an exponential tail (see, for example, [24]) and the sum of independent copies of such a variable is highly concentrated. Indeed, it is easy to see that if $\mathbb{E}\exp(|Z|/b) \leq 2$, that is, if $\|Z\|_{\psi_1} \leq b$, then $\sum_{k=1}^{\infty}(\mathbb{E}|Z|^k/b^k k!) \leq 2$. Hence, if Z_i are distributed as Z, the assumptions of Theorem 3.7 are satisfied for $M = \|Z\|_{\psi_1}$ and $v = 4n\|Z\|_{\psi_1}^2$, implying that

$$Pr\left(\left\{\left|\frac{1}{n}\sum_{i=i}^{n} Z_i\right| > u\right\}\right) \leq 2\exp\left(-cn\min\left\{\frac{u2}{\|Z\|_{\psi_1}^2}, \frac{u}{\|Z\|_{\psi_1}}\right\}\right). \qquad (3)$$

As an example, consider $Z_i = (\sum_{j=1}^{l}\xi_{i,j})^2 - l$ where, as before, $(\xi_{i,j})$ are independent, symmetric, $\{-1, 1\}$-valued random variables. It is easily verified that $\mathbb{E}\exp(Z_i/l) \leq 2$, and thus (3) is satisfied with $\|Z\|_{\psi_1} \leq l$.

Proof of Theorem 3.5. Let $m \leq N/2$, to be defined later and set P as in Lemma 3.6 for $\lambda = \frac{1}{2}$. Let $X_i = \sum_{j=1}^{n}\xi_{i,j}e_j$ and define the random vectors $Y_I = (1/m)\sum_{i\in I} X_i$. Thus, each X_i is a random point in $\{-1, 1\}^n$ and Y_I is a convex combination of points X_i out of the set $\{X_i, 1 \leq i \leq N\}$. If $I, J \in P$ and $I \neq J$ then

$$Y_I - Y_J = \frac{1}{m}\left(\sum_{i\in I\setminus J} X_i - \sum_{i\in J\setminus I} X_i\right).$$

Since the random variables $\xi_{i,j}$ are symmetric the same holds for each X_i, implying that $Y_I - Y_J$ has the same distribution as $(1/m)\sum_{i\in I\triangle J} X_i$. Thus, $(m^2/n)\cdot\|Y_I - Y_J\|_{\ell_2^n}^2$ has the same distribution as $(1/n)\sum_{j=1}^{n}(\sum_{i\in I\triangle J}\xi_{i,j})^2$.

Note that this random variable is highly concentrated. Indeed, setting $Z_j = (\sum_{i\in I\triangle J}\xi_{i,j})^2$, it is easy to see that $\|Z_j\|_{\psi_1} \leq |I \triangle J| \leq m$. Hence, by (3),

$$\begin{aligned} &Pr\left(\left\{\left|\|Y_I - Y_J\|_{\ell_2^n}^2 - \mathbb{E}\|Y_I - Y_J\|_{\ell_2^n}^2\right| > \frac{un}{m^2}\right\}\right) \\ &\quad = Pr\left(\left\{\left|\frac{1}{n}\sum_{j=1}^{n}(Z_i - \mathbb{E}Z_i)\right| > u\right\}\right) \leq 2\exp\left(-cn\min\left\{\frac{u^2}{m^2}, \frac{u}{m}\right\}\right). \quad (4)\end{aligned}$$

Since $\mathbb{E}\|Y_I - Y_J\|_{\ell_2^n}^2 = n|I \triangle J|/m^2 \geq \lambda n/m = n/2m$, then applying (4) with $u = m/4$ it follows that

$$Pr(\{\|Y_I - Y_J\|_{\ell_2^n} \leq \tfrac{1}{2}\mathbb{E}\|Y_I - Y_J\|_{\ell_2^n}\}) \leq 2\exp(-c_0 n)$$

for some absolute constant c_0. Moreover, by (4), for any $t > 0$,

$$Pr(\{\|Y_I - Y_j\|_{\ell_2^n} \geq (1+2t)\mathbb{E}\|\|Y_I - Y_j\|_{\ell_2^n}\}) \leq 2\exp(-c_0 nt),$$

and by a standard integration argument all the L_p norms of $\|Y_I - Y_J\|_{\ell_2^n}$ are equivalent to the L_1 norm with a constant depending only on p. In particular,

$$\mathbb{E}\|Y_I - Y_J\|_{\ell_2^n} \geq c(\mathbb{E}\|Y_I - Y_J\|_{\ell_2^n}^2)^{1/2} \geq c_1\sqrt{\frac{n}{m}}.$$

Therefore, with probability at least $1 - 2\exp(-c_0 n)$, $\|Y_I - Y_J\|_{\ell_2^n} \geq c_2\sqrt{n/m}$. Set $m = c_2^2 n/\varepsilon^2$ and $\kappa = c_2/\sqrt{\log 2}$. Fix $\varepsilon \geq \kappa\sqrt{\log(N/n)}$, and thus $m \leq n \leq N/2$ as required in Lemma 3.6.

Also,

$$\log|P| = (1-\lambda)m\log(c_\lambda N/m) = (m/2)\log(c'N/m)$$

and thus $2\log|P| \le c_0 n/8$. Hence, for every such ε, with probability at least $1 - \exp(-c_0 n/4)$ for every distinct $I, J \in P$, $\|Y_I - Y_J\|_{\ell_2^n} \ge \varepsilon$, implying that $K_{n,N}$ contains an ε-separated set whose cardinality satisfies that

$$\log|P| \ge \frac{m}{2}\log\left(\frac{c_0 N}{m}\right) = c_1\frac{n}{\varepsilon^2}\log\left(\frac{c_2 N\varepsilon^2}{n}\right),$$

as claimed. □

To handle scales below $\kappa\sqrt{\log(N/n)}$, we prove the following:

Lemma 3.8. *Let κ, N and n be as in Theorem* 3.5. *There exist absolute positive constants c, c_1, c_2 and c_3 for which the following holds. For any $\varepsilon \le \min\{\kappa\sqrt{\log(N/n)}, c\sqrt{n}\}$, with probability at least* $1 - \exp(-c_1 N^{1/2} n^{1/2})$,

$$\log D(K_{n,N}, \varepsilon B_2^n) \ge c_2 n \log\left(c_3\frac{\sqrt{\log(N/n)}}{\varepsilon}\right).$$

Observe that the constant κ appearing in the restriction $\varepsilon \ge \kappa\sqrt{\log(N/n)}$ is of no particular significance, and we could have chosen to use any other absolute constant. Indeed, this follows from the fact that the cardinality of an ε-separated set is monotone in the scale and since the estimates of Theorem 3.5 and of Lemma 3.8 coincide for $\varepsilon \sim \sqrt{\log(N/n)}$.

Proof. Recall that for any two convex, symmetric bodies A and B in $\mathbb{R}^n$, the covering number $N(A, B)$ satisfies that $N(A, B) \ge \mathrm{vol}(A)/\mathrm{vol}(B)$.

Hence, if we apply the volumetric estimate (1) which holds for a random $\{-1, 1\}$-polytope, it is evident that with probability $1 - \exp(-c_1 N^{1/2} n^{1/2})$,

$$D(K_{n,N}, \varepsilon B_2^n) \ge N(K_{n,N}, \varepsilon B_2^n) \ge \frac{\mathrm{vol}(K_{n,N})}{\mathrm{vol}(\varepsilon B_2^n)} \ge \left(c_2\frac{\sqrt{\log(2N/n)}}{\varepsilon}\right)^n. \qquad \square$$

Corollary 3.9. *There exist absolute positive constants c_i, $0 \le i \le 4$, and κ such that if n and N satisfy Assumption* 1, *and if we set*

$$H(\varepsilon) = c_3 n\begin{cases}\log\left(\dfrac{\sqrt{\log(2N/n)}}{\varepsilon}\right) & \text{if} \quad \varepsilon \le \kappa\sqrt{\log(N/n)},\\[2ex] \dfrac{1}{\varepsilon^2}\log\left(\dfrac{c_4 N\varepsilon^2}{n}\right) & \text{if} \quad \kappa\sqrt{\log(N/n)} \le \varepsilon \le \sqrt{n},\end{cases}$$

then with probability at least $1 - \exp(-c_0 n)$, *for any* $c_1\exp(\exp(-c_2 n)) \le \varepsilon \le \sqrt{n}$,

$$\log D(K_{n,N}, \varepsilon B_2^n) \ge H(\varepsilon).$$

Proof. By the previous results it is evident that for any fixed $0 \le \varepsilon < \sqrt{n}$, with probability at least $1 - \exp(-cn)$, $\log D(K_{n,N}, \varepsilon B_2^n) \ge H(\varepsilon)$. Fix $\varepsilon_0 = \exp(-\exp(cn))$ and $k = \exp(c'n/2)$, and let $\varepsilon_i = 2^i \varepsilon_0$ for $0 \le i \le k$. Then, with probability at least $1 - \exp(c''n)$, $\log D(K_{n,N}, \varepsilon_i B_2^n) \ge H(\varepsilon_i)$, which implies that with the same order of probability, for any $\varepsilon \in [\varepsilon_0, \sqrt{n}]$,

$$\log D(K_{n,N}, \varepsilon B_2^n) \ge cH(\varepsilon)$$

for a suitable constant c. □

We conclude by applying Theorem 3.5 to obtain a lower estimate on the Gelfand numbers of a random $K_{n,N}$. Recall that the upper estimate holds for any polytope $K_{n,N}$ and was established in [6].

Theorem 3.10. *There exist absolute positive constants c_1, c_2 and c_3 for which the following holds. For any $1 \le k \le n$ with probability at least $1 - \exp(-c_1 n)$,*

$$c_2 \min\left\{1\,,\left(\frac{\log(2N/k)}{k}\right)^{1/2}\right\} \le \frac{c_k(K_{n,N})}{\sqrt{n}} \le c_3 \min\left\{1\,,\left(\frac{\log(2N/k)}{k}\right)^{1/2}\right\}.$$

Before presenting the proof we recall the following application of a general inequality from [5].

Lemma 3.11. *There exists an absolute constant ρ such that for any symmetric convex body $K \subset \mathbb{R}^n$ and $1 \le k \le n$,*

$$\sup_{1 \le j \le k} j e_j(K) \le \rho \sup_{1 \le j \le k} j c_j(K). \tag{5}$$

Observe that in terms of entropy numbers, Theorem 3.5 states that there exist absolute constants c_1 and c_2 such that, for any $1 \le k \le n$, with probability at least $1 - \exp(-c_1 n)$, one has

$$e_k(K_{n,N}) \ge c_2 \min\left\{\sqrt{n}\,,\left(\frac{n \log(2N/k)}{k}\right)^{1/2}\right\}. \tag{6}$$

Proof of Theorem 3.10. To prove the lower estimate we can assume that $k \ge k_0 = c \log N$. Indeed, if $k < k_0$, then $c_k(K_{n,N}) \ge c_{k_0}(K_{n,N})$, while for $k = k_0$ the minimum in Theorem 3.10 is a constant. Fix k in that range and let α be a parameter larger than 1, to be defined later. From reformulation (6) of Theorem 3.5 and from (5),

$$c_3 \left(n\alpha k \log\left(\frac{2N}{\alpha k}\right)\right)^{1/2} \le \alpha k e_{\alpha k}(K_{n,N}) \le \rho \sup_{1 \le j \le \alpha k} j c_j(K_{n,N}) \tag{7}$$

for some absolute constant c_3. Clearly, one has

$$\sup_{1 \le j \le \alpha k} j c_j(K_{n,N}) \le \sup_{1 \le j < k} j c_j(K_{n,N}) + \sup_{k \le j \le \alpha k} j c_j(K_{n,N}).$$

Applying the upper bound of Theorem 2.7 for the first term on the right-hand side, it is evident that

$$\sup_{1\le j\le k} jc_j(K_{n,N}) \le c_4 \left(nk \log\left(\frac{2N}{k}\right)\right)^{1/2}. \tag{8}$$

Since for all $j \ge k,\ c_j(K_{n,N}) \le c_k(K_{n,N})$ then

$$\alpha k c_k(K_{n,N}) \ge \sup_{k\le j\le \alpha k} jc_j(K_{n,N}),$$

and combining this with (7) and (8) implies

$$c_3 \left(n\alpha k \log\left(\frac{2N}{\alpha k}\right)\right)^{1/2} - c_4\rho \left(nk \log\left(\frac{2N}{k}\right)\right)^{1/2} \le \rho\alpha k c_k(K_{n,N}).$$

To conclude, it is evident that one can choose α such that the term on the left-hand side is larger than $c_4\rho(nk\log(2N/k))^{1/2}$. □

References

1. M. Anthony, P.L. Bartlett, *Neural Network Learning*, Cambridge University Press, Cambridge, 1999.
2. I. Bárány, Z. Füredy, Approximation of the sphere by polytopes having few vertices, *Proc. Amer. Math. Soc.* **102**(3), 651–659, 1988.
3. I. Bárány and A. Pór, On 0-1 polytopes with many facets, *Adv. Math.* **161**(2001), 209–228.
4. G. Bennett, Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.* **57**, 33–45, 1962.
5. B. Carl, Inequalities of Bernstein–Jackson type and the degree of compactness of operators in Banach spaces, *Ann. Inst. Fourier* **35**, 79–118, 1985.
6. B. Carl, A. Pajor, Gelfand numbers of operators with values in a Hilbert space, *Invent. Math.* **94**, 479–504, 1988.
7. R.M. Dudley, Universal Donsker classes and metric entropy, *Ann. Probab.* **15**, 1306–1326, 1987.
8. R.M. Dudley, *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics 63, Cambridge University Press, Cambridge, 1999.
9. A. Giannopoulos, M. Hartzoulaki, Random spaces generated by vertices of the cube, *Discrete Comput. Geom.* **28**, 255–273, 2002.
10. E.D. Gluskin, Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces (in Russian), *Mat. Sb.* (N.S.) **136**(178), no. 1, 85–96, 1988; translation in *Math. USSR-Sb.* **64**(1), 85–96, 1989.
11. A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, *Adv. Math.*, to appear.
12. P. Mankiewicz, N. Tomczak-Jaegermann, Quotients of finite-dimensional Banach spaces; random phenomena, in *Handbook of the Geometry of Banach Spaces*, Vol. 2, pp. 1201–1246, North-Holland, Amsterdam, 2003.
13. S. Mendelson, G. Schechtman, The shattering dimension of sets of linear functionals, *Ann. Probab.* **32**(3A), 1746–1770, 2004.
14. S. Mendelson, R. Vershynin, Entropy and the combinatorial dimension, *Invent. Math.* **152**(1), 37–55, 2003.
15. S. Mendelson, R. Vershynin, Remarks on the geometry of coordinate projections in $\mathbb{R}^n$, *Israel J. Math.* **140**, 203–220, 2004.
16. V.D. Milman, G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces*, Lecture Notes in Mathematics 1200, Springer-Verlag, Berlin, 1986.
17. A. Pajor, *Sous espaces ℓ_1^n des espaces de Banach*, Hermann, Paris, 1985.

18. G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge University Press, Cambridge, 1989.
19. N. Sauer, On the density of families of sets, *J. Combin. Theory Ser. A*, **13**, 145–147, 1972.
20. R. Schneider, *Convex Bodies: the Brunn–Minkowski Theory*, Cambridge University Press, Cambridge, 1993.
21. S. Shelah, A combinatorial problem: stability and orders for models and theories in infinitary languages, *Pacific J. Math*. **41**, 247–261, 1972.
22. M. Talagrand, The Glivenko–Cantelli problem, *Ann. Probab.* **6**, 837–870, 1987.
23. N. Tomczak-Jaegermann, *Banach–Mazur Distances and Finite-Dimensional Operator Ideals*. Pitman Monographs and Surveys in Pure and Applied Mathematics, 38. Longman, Harlow; co-published in the United States with Wiley, New York, 1989.
24. A.W. Van der Vaart, J.A. Wellner, *Weak Convergence and Empirical Processes*, Springer-Verlag, Berlin, 1996.
25. V.N. Vapnik, A.Ya Chervonenkis, Necessary and sufficient conditions for uniform convergence of means to mathematical expectations, *Theory Probab. Appl*. **26**(3), 532–553, 1971.
26. G.M. Ziegler, Lectures on 0/1 polytopes, in *Polytopes–Combinatorics and Computation* (G. Kalai and G.M. Ziegler, eds.), pp. 1–44, DMV Seminars, Birkhäuser, Basel, 2000.

Received April 15, 2004, *and in revised form April* 26, 2005. *Online publication July* 28, 2005.