Combinatorics of random processes and sections of convex bodies

M. Rudelson* R. Vershynin†

Abstract

We find a sharp combinatorial bound for the metric entropy of sets in $\mathbb{R}^n$ and general classes of functions. This solves two basic combinatorial conjectures on the empirical processes. 1. A class of functions satisfies the uniform Central Limit Theorem if the square root of its combinatorial dimension is integrable. 2. The uniform entropy is equivalent to the combinatorial dimension under minimal regularity. Our method also constructs a nicely bounded coordinate section of a symmetric convex body in $\mathbb{R}^n$. In the operator theory, this essentially proves for all normed spaces the restricted invertibility principle of Bourgain and Tzafriri.

1 Introduction

This paper develops a sharp combinatorial method for estimating metric entropy of sets in $\mathbb{R}^n$ and, equivalently, of function classes on a probability space. A need in such estimates occurs naturally in a number of problems of analysis (functional, harmonic and approximation theory), probability, combinatorics, convex and discrete geometry, statistical learning theory, etc. Our entropy method, which evolved from the work of S.Mendelson and the second author [MV 03], is motivated by several problems in the empirical processes, asymptotic convex geometry and operator theory.

Throughout the paper, F is a class of real valued functions on some domain Ω. It is a central problem of the theory of empirical processes to determine whether the classical limit theorems hold uniformly over F. Let μ be a probability distribution

*The research supported in part by the NSF grant DMS-0245380.

†Partially supported by the NSF grant DMS-0401032 and a New Faculty Research Grant of the University of California, Davis.

on Ω and $X_1, X_2, \ldots \in \Omega$ be independent samples distributed according to a common law μ. The problem is to determine whether the sequence of real valued random variables $(f(X_i))$ obeys the central limit theorem uniformly over all $f \in F$ and over all underlying probability distributions μ, i.e. whether the random variable $\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - f(X_1))$ converges to a Gaussian random variable uniformly. With the right definition of the convergence, if that happens, F is a *uniform Donsker class*. The precise definition can be found in [LT] and [Du 99].

The pioneering work of Vapnik and Chervonenkis [VC 68, VC 71, VC 81] demonstrated that the validity of the uniform limit theorems on F is connected with the combinatorial structure of F, which is quantified by what we call the *combinatorial dimension* of F.

For a class F and $t \geq 0$, a subset σ of Ω is called t-shattered by a class F if there exists a level function h on σ such that, given any partition $\sigma = \sigma_- \cup \sigma_+$, one can find a function $f \in F$ with $f(x) \leq h(x)$ if $x \in \sigma_-$ and $f(x) \geq h(x) + t$ if $x \in \sigma_+$. The combinatorial dimension of F, denoted by $v(F, t)$, is the maximal cardinality of a set t-shattered by F. Simply speaking, $v(F, t)$ is the maximal size of a set on which F oscillates in all possible $\pm t/2$ ways around some level h.

For $\{0, 1\}$-valued function classes (classes of *sets*), the combinatorial dimension coincides with the classical *Vapnik-Chernovenkis dimension*; see [M 02] for a nice introduction to this important concept. For the integer-valued classes the notion of the combinatorial dimension goes back to 1982-83, when Pajor used it for origin symmetric classes in view of applications to the local theory of Banach spaces [Pa 82]. He proved early versions of Sauer-Shelah Lemma for sets $A \subset \{0, \ldots, p\}^n$ (see [Pa 82], [Pa 85, Lemma 4.9]). Pollard defined a similar dimension in his 1984 book on stochastic processes [Po]. Haussler also discussed this concept in his 1989 work in learning theory ([Ha], see [HL] and the references therein).

A set $A \subset \mathbb{R}^n$ can be considered as a class of functions $\{1, \ldots n\} \to \mathbb{R}$. For convex and origin symmetric sets $A \subset \mathbb{R}^n$, the combinatorial dimension $v(A, t)$ is easily seen to coincide with the maximal rank of the coordinate projection PA of A that contains the centered coordinate cube of size t. In view of this straightforward connection to convex geometry and thus to the local theory of Banach spaces, the combinatorial dimension was a central quantity in several papers of Pajor ([Pa 82], see Chapter IV of [Pa 85]). Connections of $v(F, t)$ to Gaussian processes and further applications to Banach space theory were established in the far reaching 1992 paper of M.Talagrand ([T 92], see also [T 03]). The quantity $v(F, t)$ was formally defined in 1994 by Kearns and Schapire for general classes F in their paper in learning theory [KS].

Connections between the combinatorial dimension (and its variants) with the limit theorems of probability theory have been the major theme of many papers.

For a comprehensive account of what was known about these profound connections by 1999, we refer the reader to the book of Dudley [Du 99].

Dudley proved that a class F of $\{0,1\}$-valued functions is a uniform Donsker class if and only if its combinatorial (Vapnik-Chernovenkis) dimension $v(F,1)$ is finite. This is one of the main results on the empirical processes for $\{0,1\}$ classes. The problem for general classes turned out to be much harder [T 02], [MV 03]. In the present paper we prove an optimal integral description of uniform Donsker classes in terms of the combinatorial dimension.

Theorem 1.1 *Let F be a uniformly bounded class of functions. Then*

$$\int_0^\infty \sqrt{v(F,t)}\, dt < \infty \;\Rightarrow\; F \textit{ is uniform Donsker } \;\Rightarrow\; v(F,t) = O(t^{-2}).$$

This trivially contains Dudley's theorem on the $\{0,1\}$ classes. M.Talagrand proved Theorem 1.1 with an extra factor of $\log^M(1/t)$ in the integrand and asked about the optimal value of the absolute constant exponent M [T 92], [T 02]. Talagrand's proof was based on a very involved iteration argument. In [MV 03], S.Mendelson and the second author introduced a new combinatorial idea. Their approach led to a much clearer proof, which allowed to reduce the exponent to $M = 1/2$. Theorem 1.1 removes the logarithmic factor completely, thus the optimal exponent is $M = 0$. Our argument significantly relies on the ideas originated in [MV 03] and also uses a new iteration method. The second implication of Theorem 1.1, which makes sense for $t \to 0$, is well-known ([Du 99] 10.1).

Theorem 1.1 reduces to estimating metric entropy of F by the combinatorial dimension of F. For $t > 0$, the *Koltchinskii-Pollard entropy* of F is

$$D(F,t) = \log \sup \Big(n \mid \exists f_1, \dots, f_n \in F \;\; \forall i < j \;\; \int (f_i - f_j)^2 d\mu \ge t^2 \Big)$$

where the supremum is by n and over all probability measures μ supported by the finite subsets of Ω. It is easily seen that $D(F,t)$ dominates the combinatorial dimension: $D(F,t) \gtrsim v(F,2t)$. Theorem 1.1 should then be compared to the fundamental description valid for all uniformly bounded classes:

$$\int_0^\infty \sqrt{D(F,t)}\, dt < \infty \;\Rightarrow\; F \text{ is uniform Donsker } \;\Rightarrow\; D(F,t) = O(t^{-2}). \qquad (1.1)$$

The left part of (1.1) is a strengthening of Pollard's central limit theorem and is due to Gine and Zinn (see [GZ], [Du 99] 10.3, 10.1). The right part is an observation due to Dudley ([Du 99] 10.1).

An advantage of the combinatorial description in Theorem 1.1 over the entropic description in (1.1) is that the combinatorial dimension is much easier to bound than the Koltchinskii-Pollard entropy (see [AB]). Large sets on which F oscillates in all $\pm t/2$ ways are so sound structures that their existence can be hopefully easily detected or eliminated, which leads to an estimate on the combinatorial dimension. In contrast to this, bounding Koltchinskii-Pollard entropy involves eliminating all large separated configurations $f_1, \ldots, f_n$ with respect to all probability measures μ; this can be a hard problem even on the plane (for a two-point domain Ω).

The nontrivial part of Theorem 1.1 follows from (1.1) and the central result of this paper:

Theorem 1.2 *For every class F,*

$$\int_0^\infty \sqrt{D(F,t)}\, dt \asymp \int_0^\infty \sqrt{v(F,t)}\, dt.$$

The equivalence $\asymp$ is up to an absolute constant factor C, thus $a \asymp b$ iff $a/C \le b \le Ca$.

Looking at Theorem 1.2 one naturally asks whether the Koltchinskii-Pollard entropy is pointwise equivalent to the combinatorial dimension. M. Talagrand indeed proved this for uniformly bounded classes under minimal regularity and up to a logarithmic factor. For the moment, we consider a simpler version of this regularity assumption: there exists an $a > 1$ such that

$$v(F, at) \le \frac{1}{2}\, v(F,t) \quad \text{for all } t > 0. \tag{1.2}$$

In 1992, M. Talagrand proved essentially under (1.2) that for $0 < t < 1/2$

$$c\ v(F, 2t) \le D(F,t) \le C\ v(F, ct) \log^M(1/t) \tag{1.3}$$

[T 92], see [T 87], [T 02]. Here $c > 0$ is an absolute constant and M depends only on a. The question on the value of the exponent M has been open. S. Mendelson and the second author proved (1.3) *without* the minimal regularity assumption (1.2) and with $M = 1$, which is an optimal exponent in that case. The present paper proves that *with* the minimal regularity assumption, the exponent reduces to $M = 0$, thus completely removing both the boundedness assumption and the logarithmic factor from Talagrand's inequality (1.3). As far as we know, this unexpected fact was not even conjectured.

Theorem 1.3 *Let F be a class which satisfies the minimal regularity assumption* (1.2)*. Then for all $t > 0$*

$$c\ v(F, 2t) \le D(F,t) \le C\ v(F, ct),$$

where $c > 0$ is an absolute constant and C depends only on a in (1.2)*.*

Therefore, in presence of minimal regularity, the Koltchinski-Pollard entropy and the combinatorial dimension are equivalent. Rephrasing M.Talagrand's comments from [T 02] on his inequality (1.3), Theorem 1.3 is of the type "concentration of pathology". Suppose we know that $D(F,t)$ is large. This simply means that F contains many well separated functions, but we know very little about what kind of pattern they form. The content of Theorem 1.3 is that it is possible to construct a large set σ on which not only many functions in F are well separated from each other, but on which they oscillate in *all* possible $\pm ct$ ways. We now have a very precise structure that witnesses that F is large. This result is exactly in the line of Talagrand's celebrated characterization of Glivenko-Cantelli classes [T 87], [T 96].

Theorem 1.3 remains true if one replaces the L_2 norm in the definition of the Koltchinski-Pollard entropy by the L_p norm for $1 \le p < \infty$. The extremal case $p = \infty$ is important and more difficult. The L_∞ entropy is naturally

$$D_\infty(F,t) = \log \sup \Big(n \mid \exists f_1, \ldots, f_n \in F \ \ \forall i < j \ \sup_\omega |(f_i - f_j)(\omega)| \ge t \Big).$$

Assume that F is uniformly bounded (in absolute value) by 1. Even then $D_\infty(F,t)$ can not be bounded by a function of t and $v(F,ct)$: to see this, it is enough to take for F the collection of the indicator functions of the intervals $[2^{-k-1}, 2^{-k}]$, $k \in \mathbb{N}$, in $\Omega = [0,1]$. However, if Ω is finite, it is an open question how the L_∞ entropy depends on the size of Ω. N.Alon et al. [ABCH] proved that if $|\Omega| = n$ then $D_\infty(F,t) = O(\log^2 n)$ for fixed t and $v(F,ct)$. They asked whether the exponent 2 can be reduced. We answer this by reducing 2 to any number larger than the minimal possible value 1. For every $\varepsilon \in (0,1)$,

$$D_\infty(F,t) \le Cv \, \log(n/vt) \cdot \log^\varepsilon(n/v), \quad \text{where } v = v(F, c\varepsilon t) \tag{1.4}$$

and where $C, c > 0$ are absolute constants. One can look at this estimate as a continuous asymptotic version of Sauer-Shelah Lemma. The dependence on t is optimal, but conjecturally the factor $\log^\varepsilon(n/v)$ can be removed.

The combinatorial method of this paper applies to the study of coordinate sections of a symmetric convex body K in $\mathbb{R}^n$. The average size of K is commonly measured by the so-called M-estimate, which is $M_K = \int_{S^{n-1}} \|x\|_K \, d\sigma(x)$, where σ is the normalized Lebesgue measure on the unit Euclidean sphere S^{n-1} and $\|\cdot\|_K$ is Minkowski functional of K. Passing from the average on the sphere to the Gaussian average on $\mathbb{R}^n$, Dudley's entropy integral connects the M-estimate to the integral of the metric entropy of K; then Theorem 1.2 replaces the entropy by the combinatorial dimension of K. The latter has a remarkable geometric representation, which leads to the following result. For $1 \le p \le \infty$ denote by B_p^n the unit ball of the space ℓ_p^n:

$$B_p^n = \{x \in \mathbb{R}^n \ : \ |x_1|^p + \cdots + |x_n|^p \le 1\}.$$

If M_K is large (and thus K is small "in average") then there exists a coordinate section of K contained in the normalized octahedron $D = \sqrt{n} B_1^n$. Note that the M_D is bounded by an absolute constant. In the rest of the paper, $C, C', C_1, c, c', c_1, \ldots$ will denote positive absolute constants whose values may change from line to line.

Theorem 1.4 *Let K be a symmetric convex body containing the unit Euclidean ball B_2^n, and let $M = cM_K \log^{-3/2}(2/M_K)$. Then there exists a subset σ of $\{1, \ldots, n\}$ of size $|\sigma| \ge M^2 n$, and such that*

$$M\,(K \cap \mathbb{R}^\sigma) \subseteq \sqrt{|\sigma|} B_1^\sigma. \tag{1.5}$$

Recall that the classical Dvoretzky theorem in the form of Milman guarantees, for $M = M_K$, the existence of a subspace E of dimension $\dim E \ge cM^2 n$ and such that

$$c_1 B_2^n \cap E \subseteq M(K \cap E) \subseteq c_2 B_2^n \cap E. \tag{1.6}$$

To compare the second inclusion of (1.6) to (1.5), recall that by Kashin's theorem ([K 77], [K 85], see [Pi] 6) there exists a subspace E in $\mathbb{R}^\sigma$ of dimension at least $|\sigma|/2$ such that the section $\sqrt{|\sigma|} B_1^\sigma \cap E$ is equivalent to $B_2^n \cap E$.

A reformulation of Theorem 1.4 in the operator language generalizes the restricted invertibility principle of Bourgain and Tzafriri [BT 87] to all normed spaces. Consider a linear operator $T : l_2^n \to X$ acting from the Hilbert space into arbitrary Banach space X. The "average" largeness of such an operator is measured by its ℓ-norm, defined as $\ell(T)^2 = \mathbb{E}\|Tg\|^2$, where $g = (g_1, \ldots, g_n)$ and g_i are normalized independent Gaussian random variables. We prove that if $\ell(T)$ is large then T is well invertible on some large coordinate subspace. For simplicity, we state this here for spaces of type 2 (see [LT] 9.2), which includes for example all the L_p spaces and their subspaces for $2 \le p < \infty$. For general spaces, see Section 7.

Theorem 1.5 (General Restricted Invertibility) *Let $T : l_2^n \to X$ be a linear operator with $\ell(T)^2 \ge n$, where X is a normed space of type 2. Let $\alpha = c \log^{-3/2}(2\|T\|)$. Then there exists a subset σ of $\{1, \ldots, n\}$ of size $|\sigma| \ge \alpha^2 n / \|T\|^2$ and such that*

$$\|Tx\| \ge \alpha \beta_X \|x\| \quad \text{for all } x \in \mathbb{R}^\sigma$$

where $c > 0$ is an absolute constant and $\beta_X > 0$ depends on the type 2 constant of X only.

Bourgain and Tzafriri essentially proved this restricted invertibility principle for $X = l_2^n$ (and without the logarithmic factor), in which case $\ell(T)$ equals the Hilbert-Schmidt norm of T.

The heart of our method is a result of combinatorial geometric flavor. We compare the covering number of a convex body K by a given convex body D to the number of the integer cells contained in K and its projections. This will be explained in detail in Section 2. All main results of this paper are then deduced from this principle. The basic covering result of this type and its proof occupies Section 3. First applications to covering K by ellipsoids and cubes appear in Section 4. Estimate (1.4) is also proved there. Since the proofs of Theorems 1.2 and 1.3 do not use these results, Section 4 may be skipped by a reader interested only in probabilistic applications. Section 5 deals with covering by balls of a general Lorentz space; the combinatorial dimension controls such coverings. From this we deduce in Section 6 our main results, Theorems 1.2 and 1.3. Theorem 1.2 shows in particular that *in the classical Dudley's entropy integral, the entropy can be replaced by the combinatorial dimension.* This yields a new powerful bound on Gaussian processes (see Theorem 6.5 below), which is a quantitative version of Theorem 1.1. This method is used in Section 7 to prove Theorem 1.4 on the coordinate sections of convex bodies. Theorem 1.4 is equivalently expressed in the operator language as a general principle of restricted invertibility, which implies Theorem 1.5.

ACKNOWLEDGEMENTS. The authors thank Michel Talagrand for helpful discussions.

This project started when both authors visited the Pacific Institute of Mathematical Sciences. We would like to thank PIMS for its hospitality. A significant part of the work was done when the second author was PIMS Postdoctoral Fellow at the University of Alberta. He thanks this institution and especially Nicole Tomczak-Jaegermann for support and encouragement.

2 The Method

Let K and D be convex bodies in $\mathbb{R}^n$. We are interested in the covering number $N(K, D)$, the minimal number of translates of D needed to cover K. More precisely, $N(K, D)$ is the minimal number N for which there exist points $x_1, x_2, \dots x_N$ satisfying

$$K \subseteq \bigcup_{j=1}^{N} (x_j + D).$$

Computing the covering number is a very difficult problem even in the plane [CFG]. Our main idea is to relate the covering number to the *cell content* of K, which we define as the number of the integer cells contained in all coordinate projections of

K:

$$\Sigma(K) = \sum_P \text{number of integer cells contained in } PK. \tag{2.1}$$

The sum is over all 2^n coordinate projections in $\mathbb{R}^n$, i.e. over the orthogonal projections P onto $\mathbb{R}^\sigma$ with $\sigma \subseteq \{1, \dots, n\}$. The integer cells are the unit cubes with integer vertices, i.e. the sets of the form $a + [0,1]^\sigma$, where $a \in \mathbb{Z}^\sigma$. For convenience, we include the empty set in the counting and assign value 1 to the corresponding summand.

Let D be an integer cell. To compare $N(K, D)$ to $\Sigma(K)$ on a simple example, take K to be an integer box, i.e. the product of n intervals with integer endpoints and lengths $a_i \geq 0$, $i = 1, \dots, n$. Then $N(K, D) = \prod_1^n \max(a_i, 1)$ and $\Sigma(K) = \prod_1^n (a_i + 1)$. Thus

$$2^{-n}\Sigma(K) \leq N(K, D) \leq \Sigma(K).$$

The lower bound being trivially true for any convex body K, an upper bound of this type is in general difficult to prove. This motivates the following conjecture.

Conjecture 2.1 (Covering Conjecture) *Let K be a convex body in $\mathbb{R}^n$ and D be an integer cell. Then*

$$N(K, D) \leq \Sigma(CK)^C. \tag{2.2}$$

Our main result is that the Covering Conjecture holds for a body D slightly larger that an integer cell, namely for

$$D = \Big\{ x \in \mathbb{R}^n \ : \ \frac{1}{n} \sum_1^n \exp\exp|x(i)| \leq 3 \Big\}. \tag{2.3}$$

Note that the body $5D$ contains an integer cell and the body $(5 \log\log n)^{-1} D$ is contained in an integer cell.

Theorem 2.2 *Let K be a convex body in $\mathbb{R}^n$ and D be the body* (2.3)*. Then*

$$N(K, D) \leq \Sigma(CK)^C.$$

As a useful consequence, the Covering Conjecture holds for D being an ellipsoid. This will follow by a standard factorization technique for the absolutely summing operators.

Corollary 2.3 *Let K be a convex body in $\mathbb{R}^n$ and D be an ellipsoid in $\mathbb{R}^n$ that contains an integer cell. Then*

$$N(K, D) \leq \Sigma(CK)^2.$$

As for the Covering Conjecture itself, it holds under the assumption that the covering number is exponentially large in n. More precisely, let $a > 0$ and D be an integer cell. For any $\varepsilon > 0$ and any $K \subset \mathbb{R}^n$ satisfying $N(K, D) \geq \exp(an)$, one has

$$N(K, D) \leq \Sigma(C\varepsilon^{-1}K)^M, \quad \text{where} \;\; M \leq 4\log^{\varepsilon}(1 + 1/a). \tag{2.4}$$

This result also follows from Theorem 2.2.

The usefulness of Theorem 2.2 is understood through a relation between the cell content and the combinatorial dimension. Let F be a class of real valued functions on a finite set Ω, which we identify with $\{1, \ldots, n\}$. Then we can look at F as a subset of $\mathbb{R}^n$ via the map $f \mapsto (f(i))_{i=1}^n$. For simplicity assume that F is a convex set; the general case will not be much more difficult. It is then easy to check that the combinatorial dimension $v := v(F, 1)$ equals exactly the maximal rank of a coordinate projection P in $\mathbb{R}^n$ such that PF contains a translate of the unit cube $P[0, 1]^n$. Then in the sum (2.1) for the lattice content $\Sigma(F)$, the summands with rank$P > v$ vanish. The number of nonzero summands is then at most $\sum_{k=0}^{v} \binom{n}{k}$. Every summand is clearly bounded by vol(PF), a quantity which can be easily estimated if the class F is a priori well bounded. So $\Sigma(F)$ is essentially bounded by $\sum_{k=0}^{v} \binom{n}{k}$, and is thus controlled by the combinatorial dimension v. This way, Theorem 2.2 or one of its consequences can be used to bound the entropy of F by its combinatorial dimension. Say, (2.4) implies (1.4) in this way.

In some cases, n can be removed from the bound on the entropy, thus giving an estimate independent of the size of the domain Ω. Arguably the most general situation when this happens is when F is bounded in some norm and the entropy is computed with respect to a weaker norm. The entropy of the class F with respect to a norm of a general function space X on Ω is

$$D(F, X, t) = \log \sup \Big(n \mid \exists f_1, \ldots, f_n \in F \;\; \forall i < j \;\, \|f_i - f_j\|_X \geq t \Big). \tag{2.5}$$

Koltchinskii-Pollard entropy is then $D(F, t) = \sup_\mu D(F, L_2(\mu), t)$, where the supremum is over all probability measures supported by finite sets. With the geometric representation as above,

$$D(F, X, t) = \log N_{\text{pack}}\Big(F, \frac{t}{2}\text{Ball}(X)\Big) \tag{2.6}$$

where Ball(X) denotes the unit ball of X and $N_{\text{pack}}(A, B)$ is the packing number, which is the maximal number of disjoint translates of a set $B \subseteq \mathbb{R}^n$ by vectors from a set $A \subseteq \mathbb{R}^n$. The packing and the covering numbers are easily seen to be equivalent,

$$N_{\text{pack}}(A, B) \leq N(A, B) \leq N_{\text{pack}}(A, \frac{1}{2}B). \tag{2.7}$$

To estimate $D(F, X, t)$, we have to be able to quantitatively compare the norms in the function space X an in another function space Y where F is known to be bounded. We shall consider Lorentz spaces, for which such a comparison is especially transparent. The *Lorentz space* $\Lambda_\phi = \Lambda_\phi(\Omega, \mu)$ is determined by its generating function $\phi(t)$, which is a real convex function on $[0, \infty)$, with $\phi(0) = 0$, and increasing to infinity. Then Λ_ϕ is the space of functions f on Ω such that there exists a $\lambda > 0$ for which

$$\mu\{|f/\lambda| \geq t\} \leq \frac{1}{\phi(t)} \quad \text{for all } t > 0. \tag{2.8}$$

The norm of f in Λ_ϕ is the infimum of $\lambda > 0$ satisfying (2.8). Given two Lorentz spaces Λ_ϕ and Λ_ψ, we look at their *comparison function*

$$(\phi|\psi)(t) = \sup\{\phi(s) \mid \phi(s) \geq \psi(ts)\}.$$

Under the normalization assumption $\phi(1) = \psi(1) = 1$ and a mild regularity assumption on ϕ we prove the following. If a class F is 1-bounded in Λ_ψ then for all $0 < t < 1/2$

$$D(F, \Lambda_\phi, t) \leq C\ v(F, ct) \cdot \log(\phi|\psi)(t/2). \tag{2.9}$$

An important point here is that the entropy is independent of the size of the domain Ω. To prove (2.9), we first perform a probabilistic selection, which reduces the size of Ω, and then apply Theorem 2.2, in which we replace D by a larger set $\text{Ball}(\Lambda_\phi)$.

Of particular interest are the generating functions $\phi(t) = t^p$ and $\psi(t) = t^q$ with $1 \leq p < q \leq \infty$. They define the weak L_p and L_q spaces respectively. Their comparison function is $(\phi|\psi)(t) = t^{pq/(p-q)}$. Then passing to usual L_p spaces (which is not difficult) one obtains from (2.9) the following. If F is 1-bounded in $L_q(\mu)$ then for all $0 < t < 1/2$

$$D(F, L_p(\mu), t) \leq C_{p,q}\ v(F, c_{p,q}t) \cdot \log(1/t), \tag{2.10}$$

where $C_{p,q}$ and $c_{p,q} > 0$ depend only on p and q.

First estimates of type (2.10) go back to the influential works of Vapnik and Chervonenkis. In the main combinatorial lemma of [VC 81], the volume of uniformly bounded convex class was estimated via a quantity somewhat weaker than the combinatorial dimension. Since we always have $N(K, D) \geq \text{vol}(K)/\text{vol}(D)$, the Vapnik-Chervonenkis bound is an asymptotically weaker form of (2.10) for $p = 2$ (say) and $q = \infty$. Talagrand [T 87, T 02] proved (2.10) for $p = 2, q = \infty$ up to a factor of $\log^M(1/t)$ in the right side and under minimal regularity (essentially under (1.2)). Based on the method of N.Alon et al. from [ABCH], Bartlett and Long [BL] proved (2.10) for $p = 1, q = \infty$ with an additional factor of $\log(|\Omega|/vt)$ in the right side, where $v = v(F, ct)$. The ratio $|\Omega|/v$ was removed from this factor by Bartlett, Kulkarni and Posner [BKP], thus yielding (2.10) with $\log^2(1/t)$ for $p = 1, q = \infty$. The optimal estimate (2.10) for all p and for $q = \infty$ was proved by Mendelson and the second author as the main result of [MV 03]. The present paper proves (2.10) for all p and q.

Finally, Theorems 1.2 and 1.3 are proved by iterating (2.10) with $2p = q \to \infty$ to get rid of both the logarithmic factor and any boundedness assumptions.

3 Covering by the Tower

Fix a probability space (Ω, μ). As most of our problems have a discrete nature, they essentially reduce by approximation to Ω finite and μ the uniform measure. The core difficulties arise already in this finite setting, although it took some time to fully realize this (see [T 96]). This way we shall totally ignore measurability issues.

Tower Our main covering result works for a body in $\mathbb{R}^n$ which is $\log \log n$ apart from the unit cube, while for the cube itself it remains an open problem. This body is the unit ball of the Lorentz space with generating function of the order e^{e^t}. For an extra flexibility, we shall allow a parameter $\alpha \geq 2$, generally a large number. The Lorentz space generated by the function

$$\theta(t) = \theta_\alpha(t) = e^{\alpha^t - \alpha}, \quad t \geq 1$$

is called space the *tower space* and its unit ball is called the tower. Since $\theta(1) = 1$, it does not matter how we define $\theta(t)$ for $0 < t < 1$ as long as $\theta(0) = 0$ and θ is convex; say, $\theta(t) = t$ will work. The definition of the tower space originates in the separation argument, Lemma 3.3. The proof of the main results of this paper, Theorems 1.2 and 1.3 uses an iteration procedure, which involves covering by towers with different α at each step.

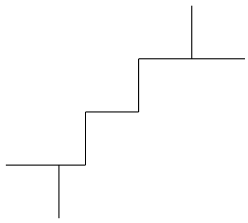
In the discrete setting, we look at Ω being $\{1, \dots, n\}$ with the uniform probability measure μ on Ω. The tower space can be realized on $\mathbb{R}^n$ by identifying a function on Ω with a point in $\mathbb{R}^n$ via the map $f \mapsto (f_i)_{i=1}^n$. The tower is then a convex symmetric body in $\mathbb{R}^n$, and we denote it by Tower^α. This body is equivalently described by (2.3),

$$c_1(\alpha) D \subseteq \text{Tower}^\alpha \subseteq c_2(\alpha) D$$

where positive $c_1(\alpha)$ and $c_2(\alpha)$ depend only on α.

Coordinate convexity We stated our results for convex bodies but not necessarily convex function classes. Convexity indeed plays very little role in our work and is replaced by a much weaker notion of *coordinate convexity*. This notion was originally motivated by problems of calculus of variations, partial differential equations and probability. The interested reader may consult the paper [M 01] and the bibliography cited there as an introduction to the subject.

One can obtain a general convex body in $\mathbb{R}^n$ by cutting off half-spaces. Similarly, a general coordinate convex body in $\mathbb{R}^n$ is obtained by cutting off octants, that is translates of the subsets of $\mathbb{R}^n$ consisting of points with fixed and nonzero signs of the coordinates. The *coordinate convex hull* of a set K in $\mathbb{R}^n$, denoted by $\mathrm{cconv}(K)$, is the minimal coordinate convex set containing K. In other words, $\mathrm{cconv}(K)$ is what remains in $\mathbb{R}^n$ after removal all octants disjoint from K. Clearly, every convex set is coordinate convex; the converse is not true, as shows the example of a cross $\{(x, y) \mid x = 0 \text{ or } y = 0\}$ in $\mathbb{R}^2$.



Example of a coordinate convex body in $\mathbb{R}^2$

Covering by the tower Let A be a nonempty set in $\mathbb{R}^n$. In contrast to what happens in classical convexity, a coordinate projection of a coordinate convex set is not necessarily coordinate convex (a pair of generic points in the plane is an example). Define the *cell content* of A as

$$\Sigma(A) = \sum_P \text{number of integer cells in } \mathrm{cconv}(PA)$$

where the sum is over all 2^n coordinate projections in $\mathbb{R}^n$, including one 0-dimensional projection, for which the summand is set to be 1. In many applications A will be a convex body, in which case $\mathrm{cconv}(PA) = PA$. The following is the main result of this section.

Theorem 3.1 *For every set F in $\mathbb{R}^n$ and $\alpha \geq 2$,*

$$N(F, \mathrm{Tower}^\alpha) \leq \Sigma(CF)^\alpha$$

where C is an absolute constant.

It is plausible that the Tower^α can be replaced by the unit cube, with α replaced by an absolute constant in the right hand side; this is a slightly stronger version of the Covering Conjecture for coordinate convex sets.

The proof of Theorem 3.1, which is a development upon [MV 03], occupies next few subsections.

Separation on one coordinate Fix a set F in $\mathbb{R}^n$ which contains more than one point. Using (2.7), we can find a finite subset $A' \subset F$ of cardinality $N(F, \text{Tower}^\alpha)$ such that no pair of points from A' lies in a common translate of $\frac{1}{2}\text{Tower}^\alpha$. Denote $A = 2A'$. Then

$$\forall x, y \in A,\ x \neq y : \quad \|x - y\|_{\text{Tower}^\alpha} \geq 1.$$

Thus for a fixed pair $x \neq y$ there exists a $t > 0$ such that $\mu\{|x - y| > t\} \geq \frac{1}{\theta(t)}$. Since $\theta(t) < 1$ for $t < 1$, we necessarily have $t \geq 1$, hence

$$\exists t > 0 : \quad \mu\{|x - y| > t\} \geq \frac{1}{\tilde{\theta_\alpha}(t)}$$

where

$$\tilde{\theta_\alpha}(t) = e^{\alpha^t - \alpha}, \qquad t \geq 0.$$

By Chebychev's inequality,

$$\mathbb{E}_i\, \tilde{\theta_\alpha}(|x(i) - y(i)|) \geq 1,$$

where $\mathbb{E}_i$ is the expectation according to the uniform distribution of the coordinate i in $\{1, \ldots, n\}$. Let x and y be random points drawn from A independently and according to the uniform distribution on A. Then $x \neq y$ with probability $1 - |A|^{-1} \geq \frac{1}{2}$, and taking the expectation with respect to x and y, we obtain

$$\mathbb{E}_{x,y}\, \mathbb{E}_i\, \tilde{\theta_\alpha}(|x(i) - y(i)|) \geq \frac{1}{2}.$$

Changing the order of the expectation, we find a realization of the random coordinate i for which

$$\mathbb{E}_{x,y}\, \tilde{\theta_\alpha}(|x(i) - y(i)|) \geq \frac{1}{2}. \tag{3.1}$$

Fix this realization.

Recall that a median of a real valued random variable ξ is a number M satisfying $\mathbb{P}(\xi \leq M) \geq 1/2$ and $\mathbb{P}(\xi \geq M) \geq 1/2$. Unlike the expectation, the median may be not uniquely defined. We can replace $y(i)$ in (3.1) by a median of $x(i)$ using the following standard observation.

Lemma 3.2 *Let ϕ be a convex and nondecreasing function on $[0,\infty)$. Let X and Y be identically distributed random variables. Then*

$$\inf_a \mathbb{E}\,\phi(|X-a|) \le \mathbb{E}\,\phi(|X-Y|) \le \inf_a \mathbb{E}\,\phi(2|X-a|).$$

Proof. The first inequality follows from Jensen's inequality with $a = \mathbb{E}X = \mathbb{E}Y$. For the second one, the assumptions on ϕ imply through the triangle and Jensen's inequalities that for every a

$$\phi(|X-Y|) \le \phi(|X-a| + |Y-a|) \le \frac{1}{2}\phi(2|X-a|) + \frac{1}{2}\phi(2|Y-a|).$$

Taking the expectations on both sides completes the proof. ■

Denote by M a median of $x(i)$ over $x \in A$. We conclude that

$$\mathbb{E}_x\, \tilde{\theta_\alpha}(2|x(i)-M|) \ge \frac{1}{2}. \tag{3.2}$$

Lemma 3.3 (Separation Lemma) *Let X be a random variable with median M. Assume that for every real a*

$$\mathbb{P}\{X \le a\}^{1/\alpha} + \mathbb{P}\{X > a+1\}^{1/\alpha} \le 1.$$

Then

$$\mathbb{E}\,\tilde{\theta_\alpha}(c|X-M|) < \frac{1}{2}.$$

In particular, the conclusion implies that the tower norm of the random variable $X - M$ is bounded by an absolute constant.

Proof. One can assume that $M = 0$. With the notation $p(a) = \mathbb{P}\{X > a\}$, the assumption of the lemma implies that for every a

$$(1-p(a)) + (p(a+1))^{1/\alpha} \le (1-p(a))^{1/\alpha} + (p(a+1))^{1/\alpha} \le 1,$$

hence

$$p(a+1) \le p(a)^{1/\alpha}, \quad a \in \mathbb{R}.$$

Applying this estimate successively and using $p(0) = 1 - \mathbb{P}(x \le 0) \le \frac{1}{2}$, we obtain $p(k) \le 2^{-\alpha^k}, \quad k \in \mathbb{N}$. Then for every real number $a \ge 2$

$$p(a) \le p([a]) \le 2^{-\alpha^{[a]}} \le 2^{-\alpha^{a-1}} \le 2^{-\alpha^{a/2}}.$$

Repeating this argument for $-X$, we conclude that

$$\mathbb{P}\{|X| > a\} \le 2^{1-\alpha^{a/2}}, \quad a \ge 2.$$

Then

$$\mathbb{P}\{e^{\alpha^{c|X|}} > s\} \le 2^{1-(\log s)^{1/2c}} \le 2s^{-\alpha^{1-2c}}, \quad s \ge e^{\alpha^{2c}}.$$

Integrating by parts and using this tail estimate, we have

$$\mathbb{E}\,\tilde{\theta_\alpha}(c|X|) = e^{-\alpha}\mathbb{E}e^{\alpha^{c|X|}} \le e^{-\alpha}\Big[e^{\alpha^{2c}} + \int_{e^{\alpha^{2c}}}^{\infty} 2s^{-\alpha^{1-2c}}\,ds\Big]$$
$$= e^{-\alpha+\alpha^{2c}} + 2(\alpha^{1-2c}-1)^{-1}e^{-2\alpha+\alpha^{2c}} =: h(\alpha, c).$$

For a fixed $c \le 1/4$, the function $h(\alpha, c)$ decreases as a function of α on $[2, \infty)$, and $h(2, 0) = e^{-1} + 2e^{-3} \approx 0.47 < \frac{1}{2}$. Hence for a suitable choice of the absolute constant $c > 0$,

$$h(\alpha, c) \le h(2, c) < \frac{1}{2}$$

because $\alpha \ge 2$. This completes the proof. ■

Applying the Separation Lemma to the random variable $\frac{2}{c}x(i)$ together with (3.2), we find an $a \in \mathbb{R}$ so that

$$\mu\{x(i) \le a\}^{1/\alpha} + \mu\{x(i) > a + c\}^{1/\alpha} > 1,$$

where μ is the uniform measure on A. Equivalently, for the subsets A_- and A_+ of A defined as

$$A_- = \{x \,:\, x(i) \le a\}, \quad A_+ = \{x \,:\, x(i) > a + c\} \tag{3.3}$$

we have

$$|A_-|^{1/\alpha} + |A_+|^{1/\alpha} > |A|^{1/\alpha}. \tag{3.4}$$

Here $|A|$ denotes the cardinality of the set A.

Separating tree This and the next step are versions of corresponding steps of [MV 03], where they were written in terms of function classes. Continuing the process of separation for each A_- and A_+, we construct a *separating tree* of subsets of A.

A tree of nonempty subsets of a set A is a finite collection T of subsets of A such that every two elements in T are either disjoint or one contains the other. A *son* of an element $B \in T$ is a maximal (with respect to inclusion) proper subset of B which belongs to T. An element of with no sons is called a *leaf*, an element which is not a son of any other element is called a *root*.

Definition 3.4 *Let A be a class of functions on Ω and $t > 0$. A t-*separating tree *T* of *A is a tree of subsets of A whose only root is A and such that every element $B \in T$ which is not a leaf has exactly two sons B_+ and B_- and, for some coordinate $i \in \Omega$,*

$$f(i) \geq g(i) + t \quad \textit{for all } f \in B_+,\ g \in B_-.$$

If $|A_-| > 1$, we can repeat the separation on one coordinate for A_- (note that this coordinate may be different from i). The same applies to A_+. Continuing this process of separation until all the resulting sets are singletons, we arrive at

Lemma 3.5 *Let $A \subset \mathbb{R}^n$ be a finite set whose points are* 1*-separated in* Tower^α*-norm. Then there exists a c-separating tree of A with at least $|A|^{1/\alpha}$ leaves.*

This separating tree improves in a sense the set A which was already separated. Of course, the leaves in this tree are c-separated in the L_∞-norm, but the tree also shows some pattern in the coordinates on which they are separated. This will be used in the next section where we further improve the separation of A by constructing in it many copies of a discrete cube (on different subsets of coordinates).

However note that the assumption on A, that it is separated in the tower norm, is stronger than being separated in the L_∞-norm.

Proof. We proceed by induction on the cardinality of A. The claim is trivially true for singletons. Assume that $|A| > 1$ and that the claim holds for all sets of cardinality smaller than $|A|$. By the separation procedure described above, we can find two subsets A_- and A_+ satisfying (3.3) and (3.4). The strict inequality in (3.4) implies that the cardinalities of both sets is strictly smaller than $|A|$. By the induction hypothesis, both A_- and A_+ have c-separating trees T_- and T_+ with at least $|A_-|^{1/\alpha}$ and $|A_+|^{1/\alpha}$ leaves respectively.

Now glue the trees T_- and T_+ into one tree T of subsets of A by declaring A the root of T and A_- and A_+ the sons of A. By (3.3), $f(i) \geq g(i) + c$ for all $f \in A_+$, $g \in A_-$. Therefore T is a c-separating tree of A. The number of leaves in T is the sum of the number of leaves of T_- and T_+, which is at least $|A_-|^{1/\alpha} + |A_+|^{1/\alpha} > |A|^{1/\alpha}$ by (3.4). This proves the lemma. ∎

Coordinate convexity and counting cells Recall that $|A| = N(F, \mathrm{Tower}^\alpha)$. We shall prove the following fact which, together with Lemma 3.5, finishes the proof.

Lemma 3.6 *Let A be a set in $\mathbb{R}^n$, and T be a* 2*-separating tree of A. Then*

$$\textit{Number of leaves in } T \leq \Sigma(A).$$

The value 2 is exact here. For example, the open cube $A = (-1, 1)^n$ has $\Sigma(A) = 1$, because A contains no integer cells. However, for every $\varepsilon > 0$ one easily constructs a $(2-\varepsilon)$-separating tree of A with 2^n leaves.

We ask what it means for a cell to be contained in the coordinate convex hull of a set. A cell $\mathcal{C}$ in $\mathbb{R}^n$ defines 2^n octants in a natural way. Let $\theta \in \{-1, 1\}^n$ be a choice of signs. A closed octant with the vertex $z \in \mathbb{R}^n$ is the set

$$\mathcal{O}_\theta(z) = \{x = (x_1, \dots x_n) \in \mathbb{R}^n \mid (x_i - z_i) \cdot \theta_i \geq 0 \quad \text{for } i = 1, \dots n\}.$$

The octants generated by a cell are those who have only one common point with it (a vertex).

Lemma 3.7 *Let A be a set in $\mathbb{R}^n$ and $\mathcal{C}$ be a cell of $\mathbb{Z}^n$. Then $\mathcal{C} \subset \operatorname{cconv}(A)$ if and only if A intersects all the octants generated by $\mathcal{C}$.*

The proof is straightforward and we omit it. ∎

Proof of Lemma 3.6. It will suffice to prove that

$$\text{if } A_- \text{ and } A_+ \text{ are the sons of } A, \text{ then } \Sigma(A_-) + \Sigma(A_+) \leq \Sigma(A). \qquad (3.5)$$

Indeed, assuming that (3.5) one can complete the proof by induction on the cardinality of A as follows. The lemma is trivially true for singletons. Assume that $|A| > 1$ and that the lemma holds for all sets of cardinality smaller than $|A|$. Let A_- and A_+ be the sons of A. Define T_- to be the colection of sets from T that belong to A_-; then T_- is a separating tree of A_-. Do similarly for T_+. Since both A_- and A_+ have cardinalities smaller than $|A|$, the induction hypothesis applies to them. Hence by (3.5) we have

$$\begin{aligned}\Sigma(A) \geq \Sigma(A_-) + \Sigma(A_+) &\geq (\text{number of leaves in } T_-) + (\text{number of leaves in } T_+) \\ &= \text{number of leaves in } T.\end{aligned}$$

This proves the lemma, so the only remaining thing is to prove (3.5).

In the proof of (3.5), when it creates no confusion, we will denote by $\Sigma(A)$ not only the cardinality, but also the set of all pairs $(P, \mathcal{C})$ for which $\mathcal{C} \subset \operatorname{cconv}(PA)$. For this to be consistent, we introduce a 0-dimensional cell $\emptyset$, and always assume that the 0-dimensional projection along with the empty cell are in $\Sigma(A)$ provided A is nonempty.

Clearly, $\Sigma(A_-) \cup \Sigma(A_+) \subseteq \Sigma(A)$. To complete the proof, it will be enough to construct an injective mapping Φ from $\Sigma(A_-) \cap \Sigma(A_+)$ into $\Sigma(A) \setminus (\Sigma(A_-) \cup \Sigma(A_+))$.

We will do this by gluing identical cells from $\Sigma(A_-) \cap \Sigma(A_-)$ into a larger cell; this idea goes back to [ABCH].

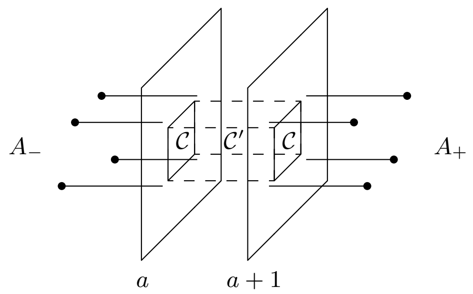
Fix a pair $(P, \mathcal{C}) \in \Sigma(A_-) \cap \Sigma(A_+)$. Without loss of generality, we may assume that A_- and A_+ are 2-separated on the first coordinate. Then there exists an integer a such that

$$x(1) \leq a \text{ for } x \in A_-, \quad x(1) \geq a+1 \text{ for } x \in A_+. \tag{3.6}$$

The coordinate projection P must annihilate the first coordinate, otherwise (3.6) would imply that the sets PA_- and PA_+ are disjoint, which would contradict to our assumption that their coordinate convex hulls both contain the cell $\mathcal{C}$.

Trivial case: rank$P = 0$. In this case, let P' be the coordinate projection that annihilates all the coordinates except the first. Since both A_- and A_+ are nonempty, $P'A$ contains points for which $x(1) \leq a$ and $x(1) \geq a+1$. Hence $\mathrm{cconv}(P'A)$ contains the one-dimensional cell $\mathcal{C}' = [a, a+1]$. So, we can define the action of Φ on the trivial pair as $\Phi : (P, \emptyset) \mapsto (P', \mathcal{C}')$.

Nontrivial case: rank$P > 0$. Without loss of generality we may assume that P retains the coordinates $\{2, 3, \ldots, k\}$ with some $2 \leq k \leq n$, and annihilates the others. Let P' be the coordinate projection onto $\mathbb{R}^k$, so $\mathcal{C}' = [a, a+1] \times \mathcal{C}$ is a cell in $\mathbb{R}^k$. We claim that $(P', \mathcal{C}') \in \Sigma(A)$. By the assumption, the cell $\mathcal{C}$ lies in both $\mathrm{cconv}(PA_-)$ and $\mathrm{cconv}(PA_+)$. In light of Lemma 3.7, PA_- and PA_+ each intersect all the octants generated by $\mathcal{C}$, and we need to show that PA' intersects any octant $\mathcal{O}'$ generated by $\mathcal{C}'$. This octant must be of the form either $\mathcal{O}' = \{x \in \mathbb{R}^k \ : \ x(1) \leq a, Px \in \mathcal{O}\}$ or $\mathcal{O}' = \{x \in \mathbb{R}^k \ : \ x(1) \geq a+1, Px \in \mathcal{O}\}$, where $\mathcal{O}$ is some octant generated by the cell $\mathcal{C}$. Assume the second option holds. Pick a point $z \in A_+$ such that $Pz \in PA_+ \cap \mathcal{O}$. Then $P'z(1) = z(1) \geq a+1$, so $P'z \in P'A_+ \cap \mathcal{O}'$. A similar argument (with A_-) works if $\mathcal{O}$ is of the first form. This proves the claim, and we again define the action of Φ as $\Phi : (P, \mathcal{C}) \mapsto (P', \mathcal{C}')$.



Nontrivial case: Gluing two copies of $\mathcal{C}$ into a larger cell $\mathcal{C}'$

To check that the range of Φ is disjoint from both $\Sigma(A_-)$ and $\Sigma(A_+)$, assume that the pair $(P', \mathcal{C}')$ constructed above is in $\Sigma(A_-)$. This means that $\mathcal{C}'$ lies in cconv(QA_-) for some coordinate projection Q. This projection must retain the first coordinate because the cell $\mathcal{C}'$ is non-degenerating on the first coordinate by its construction. Therefore, since $x(1) \le a$ for all $x \in A_-$, the same must hold for all $x \in Q(A_-)$, and hence also for all $x \in \mathrm{cconv}(QA_-)$. On the other hand, there clearly exist points in $\mathcal{C}'$ with $x(1) = a+1 > a$. Hence $\mathcal{C}'$ can not lie in cconv(QA_-). A similar argument works for A_+. Therefore the range of Φ is as claimed.

Finally, Φ is trivially injective because the map $\mathcal{C} \mapsto \mathcal{C}'$ is injective. ■

Theorem 3.1 follows from Lemma 3.5 and Lemma 3.6.

Remark. The proof does not use the fact that the probability measure on $\Omega = \{1, \dots, n\}$, underlying the tower space, is uniform. In fact, Theorem 3.1 holds for any probability measure on $\{1, \dots, n\}$. This will help us in next section.

4 Covering by Ellipsoids and Cubes

The Covering Conjecture holds if we cover by ellipsoids containing the unit cube rather by the unit cube itself. This nontrivial fact is a consequence of Theorem 3.1.

Theorem 4.1 *Let A be a set in $\mathbb{R}^n$ and D be an ellipsoid containing the cube $[0,1]^n$. Then*

$$N(A, D) \le \Sigma(CA)^2$$

where C is an absolute constant.

This result will be used in Section 7 to find nice sections of convex bodies.

Proof. Translating the ellipsoid D, we can assume that $2D$ contains the cube $[-1,1]^n$, which is the unit ball of the space l_∞^n. Call X the normed space $(\mathbb{R}^n, \|\cdot\|_{2D})$. Then X is isometric to l_2^n. Let $T : l_\infty^n \to X$ be the formal identity map and $S : X \to l_2^n$ be an isometry. Finally, define $u = ST : l_\infty^n \to l_2^n$ and note that $\|u\| \le 1$. Recall that every linear operator $u : l_\infty^n \to l_2^n$ is 2-summing and its 2-summing norm $\pi_2(u)$ satisfies $\pi_2(u) \le \sqrt{\pi/2}\ \|u\|$, see [TJ] Corollary 10.10. Thus $\pi_2(u) \le \sqrt{\pi/2}$. By Pietsch's factorization theorem (see [TJ] Theorem 9.3) there exists a probability measure μ on $\Omega = \{1, \dots, n\}$ such that for all $x \in \mathbb{R}^n$

$$\|ux\| \le \sqrt{\pi/2}\ \|x\|_{L_2(\Omega,\mu)}.$$

Since $\|ux\| = \|S^{-1}ux\|_X = \|Tx\|_X = \|x\|_X$, we have

$$\frac{1}{\sqrt{\pi/2}} \|x\|_X \le \|x\|_{L_2(\Omega,\mu)}. \tag{4.1}$$

On the other hand, the norm of the Lorentz space generated by $\theta_2(t) = e^{2^t-2}$ clearly dominates the L_2 norm: for every $x \in \mathbb{R}^n$,

$$\|x\|_{L_2(\Omega,\mu)} \leq C\|x\|_{\Lambda_{\theta_2}(\Omega,\mu)} \tag{4.2}$$

where C is an absolute constant. Denoting by $\text{Tower}^2(\mu)$ the unit ball of the norm in the right hand side of (4.2), we conclude from (4.1) and (4.2) that

$$\text{Tower}^2(\mu) \subseteq C'D$$

where C' is an absolute constant. Then by Theorem 3.1 and the remark after its proof,

$$N(A, D) \leq N(C'A, \text{Tower}^2(\mu)) \leq \Sigma(C''A)^2$$

where C'' is an absolute constant. ■

Next theorem is a partial positive solution to the Covering Conjecture itself. We prove the conjecture with a mildly growing exponent.

Theorem 4.2 *Let A be a set in $\mathbb{R}^n$ and $\varepsilon > 0$. Then for the integer cell $Q = [0, 1]^n$*

$$N(A, Q) \leq \Sigma(C\varepsilon^{-1}A)^M$$

with $M = 4\log^\varepsilon(e + n/\log N(A, Q))$, and where C is an absolute constant.

In particular, this proves the Covering Conjecture in case when the covering number is exponential in n: if $N(A, Q) \geq \exp(\lambda n)$, $\lambda < 1/2$, then $M \leq C\log^\varepsilon(1/\lambda)$.

For the proof of the theorem, we first cover A by towers, and then towers by cubes. Formally,

$$\begin{aligned} N(A, Q) &\leq N(A, \varepsilon\text{Tower}^\alpha)\ N(\varepsilon\text{Tower}^\alpha, Q) \\ &= N(\varepsilon^{-1}A, \text{Tower}^\alpha)\ N(\text{Tower}^\alpha, \varepsilon^{-1}Q). \end{aligned} \tag{4.3}$$

Lemma 4.3 *For every $t \geq 4$,*

$$N(\text{Tower}^\alpha, tQ) \leq \exp(Ce^{-\frac{1}{4}\alpha^{t/2}}n)$$

where C is an absolute constant.

Proof. We count the integer points in the tower. For $x \in \mathbb{R}^n$, define a point $x' \in \mathbb{Z}^n$ by $x'(i) = \text{sign}(x(i))[x(i)]$. Every point $x \in \text{Tower}^\alpha$ is covered by the cube $x' + [-1,1]^n$, so

$$N = N(\text{Tower}^\alpha, tQ) = N(2t^{-1}\text{Tower}^\alpha, 2Q) \le |\{x' \in \mathbb{Z}^n \mid x \in 2t^{-1}\text{Tower}^\alpha\}| \\ \le |2t^{-1}\text{Tower}^\alpha \cap \mathbb{Z}^n|.$$

For every $x \in 2t^{-1}\text{Tower}^\alpha \cap \mathbb{Z}^n$,

$$|\{i : |x(i)| = j\}| \le e^{-\alpha^{tj/2}+\alpha} n =: k_j, \quad j \in \mathbb{N}.$$

Let J be the largest number j such that $k_j \ge 1$. Then

$$N \le \prod_{j=1}^{J} \binom{n}{k_j} 2^{k_j},$$

as for every j there are at most $\binom{n}{k_j}$ ways to choose the the level set $\{i : |x(i)| = j\}$, and at most 2^{k_j} ways to choose signs of $x(i)$.

Let $\beta_j = k_j/n$. Since $\alpha \ge 2$ and $t \ge 2$, $\beta_j < 1/4$. Then $\binom{n}{k_j} \le \big(e/\beta_j\big)^{\beta_j n} \le \exp(C\beta_j^{1/2} n)$. Hence

$$N \le \exp\Big(C_1 \sum_{j=1}^{J} \beta_j^{1/2} n\Big) \le \exp(C_2 \beta_1^{1/2} n) \le \exp(C_2 e^{-\frac{1}{4}\alpha^{t/2}} n).$$

This completes the proof. ■

Proof of Theorem 4.2. We can assume that $0 < \varepsilon < c$ where $c > 0$ is any absolute constant. We estimate the second factor in (4.3) by Lemma 4.3. With $\alpha = M/2$,

$$N(\text{Tower}^\alpha, \varepsilon^{-1}Q) \le \exp\Big[C\Big(e + \frac{n}{\log N(A,Q)}\Big)^{-2^{1/2\varepsilon}/4} n\Big] \\ \le \exp\Big[Ce^{-2^{1/2\varepsilon}/4+1}\Big(\frac{n}{\log N(A,Q)}\Big)^{-1} n\Big] \\ \le N(A,Q)^{1/2}.$$

Then (4.3) and Theorem 3.1 imply that

$$N(A,Q) \le N(\varepsilon^{-1}A, \text{Tower}^{M/2})^2 \le \Sigma(c\varepsilon^{-1}A)^M.$$

The proof is complete. ■

Theorem 4.2 applies to a combinatorial problem studied by N.Alon et al. [ABCH].

Theorem 4.4 *Let F be a class of functions on an n-point set Ω with the uniform probability measure μ. Assume F is 1-bounded in $L_1(\Omega, \mu)$. Then for $0 < \varepsilon < 1$ and for $0 < t < 1/2$*

$$D_\infty(F, t) \le Cv \, \log(n/vt) \cdot \log^\varepsilon(2n/v) \tag{4.4}$$

where $v = v(F, c\varepsilon t)$.

N.Alon et al. [ABCH] proved under a somewhat stronger assumption (F is 1-bounded in L_∞) that

$$D_\infty(F, t) \le Cv \, \log(n/vt) \cdot \log(n/t^2), \quad \text{where } v = v(F, ct). \tag{4.5}$$

Thus $D_\infty(F, t) = O(\log^2 n)$. It was asked in [ABCH] whether the exponent 2 can be reduced to some constant between 1 and 2. Theorem 4.4 answers this in positive. It remains open whether the exponent can be made 1. A partial case of Theorem 4.4, for $\varepsilon = 2$ and for uniformly bounded classes, was proved in [MV 02].

It is important that, unlike in (4.5), the size of the domain n appears in (4.4) always in the ratio n/v. Assume, for example, that one knows *a priori* that the entropy is large: for some constant $0 < a < 1/2$

$$D_\infty(F, t) \ge an.$$

Then by (4.4) we have $an \le Cv \, \log(n/vt) \cdot \log^\varepsilon(2n/v)$. Dividing by n and solving for n/v, we get

$$n/v \le \frac{C}{a} \Big[\log\Big(\frac{1}{t}\Big) \log^\varepsilon \Big(\frac{1}{a} \log\frac{1}{t}\Big) + \log^{1+\varepsilon}\Big(\frac{1}{a}\Big) \Big]$$

and putting this back into (4.4) we obtain

$$D_\infty(F, t) \le Cv \, \log\Big(\frac{1}{at}\Big) \cdot \log^\varepsilon \Big(\frac{1}{a} \log\frac{1}{t}\Big).$$

We see that n, the size of the domain Ω, disappeared from the entropy estimate. Such domain-free bounds, to which we shall return in the next section, are possible only because n enters into the entropy estimate (4.4) in the ratio n/v.

To prove Theorem 4.4, we identify the n-point domain Ω with $\{1, \dots, n\}$ and realize the class of functions F as a subset of $\mathbb{R}^n$ via the map $f \mapsto (f(i))_{i=1}^n$. The geometric meaning of the combinatorial dimension of F is then the following.

Definition 4.5 *The* combinatorial dimension $v(A)$ *of a set* A *in* $\mathbb{R}^n$ *is the maximal rank of a coordinate projection* P *in* $\mathbb{R}^n$ *so that* $\operatorname{cconv}(PA)$ *contains an integer cell.*

This agrees with the classical Vapnik-Chernovenkis definition for sets $A \subseteq \{0,1\}^n$, for which $v(A)$ is defined as the maximal rank of a coordinate projection P such that $PA = P(\{0,1\}^n)$.

Lemma 4.6 $v(F,1) = v(F)$, *where* F *is treated as a function class in the left hand side and as a subset of* $\mathbb{R}^n$ *in the right hand side.*

Proof. By the definition, $v(F,1)$ is the maximal cardinality of a subset σ of $\{1,\dots,n\}$ which is 1-shattered by F. Being 1-shattered means that there exists a point $h \in \mathbb{R}^n$ such that for every partition $\sigma = \sigma_- \cup \sigma_+$ one can find a pont $f \in F$ with $f(i) \le h(i)$ if $i \in \sigma_-$ and $f(i) \ge h(i)+1$ if $i \in \sigma_+$. This means exactly that $P_\sigma F$ intersects each octant generated by the cell $\mathcal{C} = h + [0,1]^\sigma$, where P_σ denotes the coordinate projection in $\mathbb{R}^n$ onto $\mathbb{R}^\sigma$. By Lemma 3.7 this means that $\mathcal{C} \subset \operatorname{cconv}(PF)$. Hence $v(F,1) = v(F)$. ■

For a further use, we will prove Theorem 4.4 under a weaker assumption, namely that F is 1-bounded in $L_p(\mu)$ for some $0 < p < \infty$. When F is realized as a set in $\mathbb{R}^n$, this assumption means that F is a subset of the unit ball of L_p^n, which is

$$\operatorname{Ball}(L_p^n) = \Big\{x \in \mathbb{R}^n \ : \ \sum_1^n |x(i)|^p \le n\Big\}.$$

We will apply to F the covering Theorem 4.2 and then estimate $\Sigma(F)$ as follows.

Lemma 4.7 *Let* A *be a subset of* $a \cdot \operatorname{Ball}(L_p^n)$ *for some* $a \ge 1$ *and* $0 < p \le \infty$. *Then*

$$\Sigma(A) \le \Big(\frac{C_1(p)an}{v}\Big)^{C_2(p)v}$$

where $v = v(A)$, $C_1(p) = C(1 + \frac{1}{\sqrt{p}})$ *and* $C_2(p) = 1 + \frac{1}{p}$.

Proof. We look at

$$\Sigma(A) = \sum_P \text{number of integer cells in } \operatorname{cconv}(PA)$$

and notice that by Lemma 4.6, $\operatorname{rank}P \le v(A) = v$ for all P in this sum. Since the number of integer cells in a set is always bounded by its volume,

$$\Sigma(A) \le \sum_{\operatorname{rank}P \le v} \operatorname{vol}(\operatorname{cconv}(PA)) \le \sum_{\operatorname{rank}P \le v} \operatorname{vol}\Big(P(a \cdot \operatorname{Ball}(L_p^n))\Big)$$

where the volumes are considered in the corresponding subspaces $P(\mathbb{R}^n)$. By the symmetry of L_p^n, the summands with the same rank$P$ in the last sum are equal. Then the sum equals

$$1 + \sum_{k=1}^{v} \binom{n}{k} a^k \operatorname{vol}_k \Big(P_k(\operatorname{Ball}(L_p^n)) \Big) \tag{4.6}$$

where P_k denotes the coordinate projection in $\mathbb{R}^n$ onto $\mathbb{R}^k$. Note that $P_k(\operatorname{Ball}(L_p^n)) = (n/k)^{1/p} \operatorname{Ball}(L_p^k)$ and recall that $\operatorname{vol}(\operatorname{Ball}(L_p^k)) \le C_1(p)^k$, see [Pi] (1.18). Then the volumes in (4.6) are bounded by $(n/k)^{k/p} C_1(p)^k \le (C_1(p)n/k)^{C_2(p)k}$. The binomial coefficients in (4.6) are estimated via Stirling's formula as $\binom{n}{k} \le (en/k)^k$. Then (4.6) is bounded by

$$1 + \sum_{k=1}^{v} \Big(\frac{en}{k}\Big)^k a^k \Big(\frac{n}{k}\Big)^{k/p} C_1(p)^k \le \Big(\frac{C \cdot C_1(p) an}{v}\Big)^{C_2(p)v}.$$

This completes the proof. ■

Proof of Theorem 4.4. Viewing F as a set in $\mathbb{R}^n$, we notice from (2.6) and (2.7) that

$$D_\infty(F, t) \le \log N(F, 2tQ) \le D_\infty(F, t/2) \tag{4.7}$$

where $Q = [0,1]^n$. Therefore it is enough to estimate $N = N(F, 2tQ)$. We apply successively the covering Theorem 4.2 and Lemma 4.7 with $p = 1$:

$$N = N\big(\frac{1}{2t} F, Q\big) \le \Sigma\Big(\frac{C}{\varepsilon t} F\Big)^M \le \Big(\frac{Cn}{\varepsilon t v}\Big)^{CMv} \tag{4.8}$$

where $v = v(\frac{c}{\varepsilon t} F) = v(F, \frac{\varepsilon t}{c})$ and $M = 4 \log^\varepsilon(e + n/\log N)$. Define the number $a > 0$ by $N = \exp(an)$. Then $M = 4\log^\varepsilon(e + \frac{1}{a})$ and taking logarithms in (4.8) we have $an \le CMv \log(\frac{Cn}{\varepsilon v t})$. Dividing by Mn, we obtain

$$\frac{a}{\log^\varepsilon(e + \frac{1}{a})} \le \frac{Cv}{n} \log\Big(\frac{Cn}{\varepsilon v t}\Big).$$

This implies

$$a \le \frac{Cv}{n} \log\Big(\frac{Cn}{\varepsilon v t}\Big) \log^\varepsilon\Big(\frac{Cn}{v} / \log\Big(\frac{Cn}{\varepsilon v t}\Big)\Big) \le \frac{Cv}{n} \log\Big(\frac{Cn}{\varepsilon v t}\Big) \log^\varepsilon\Big(\frac{Cn}{v}\Big)$$

and multiplying by n we obtain

$$\log N \le Cv \log(Cn/v\varepsilon t) \cdot \log^\varepsilon(Cn/v). \tag{4.9}$$

It remains to remove ε from the denominator by a routine argument.

Consider the function

$$\phi(\varepsilon) = \log^{\varepsilon}(Cn/v), \quad \text{where } v = v(\varepsilon) \text{ as before.}$$

As ε decreases to zero, $v(\varepsilon)$ increases, thus $\phi(\varepsilon)$ decreases to 1. Define ε_0 so that $\phi(\varepsilon_0) = e$.

Case 1. Assume that $\varepsilon \geq \varepsilon_0$. Then $\phi(\varepsilon) \geq e$, thus $\varepsilon \geq 1/\log\log(Cn/v)$, so $Cn/v\varepsilon t \leq (Cn/vt)^2$. Using this in (4.9) we obtain

$$\log N \leq Cv \, \log(Cn/vt) \cdot \log^{\varepsilon}(Cn/v). \tag{4.10}$$

Case 2. Let $\varepsilon < \varepsilon_0$. Then $\phi(\varepsilon) \leq e$, so by (4.9),

$$\log N \leq Cv(\varepsilon_0) \, \log(Cn/v(\varepsilon_0)\varepsilon_0 t) \cdot e. \tag{4.11}$$

As in case 1, we have $Cn/v(\varepsilon_0)\varepsilon_0 t \leq (Cn/v(\varepsilon_0)t)^2$. Using this in (4.11), we obtain

$$\log N \leq C' v(\varepsilon_0) \, \log(Cn/v(\varepsilon_0)t) \leq C' v \log(Cn/vt),$$

because $v(\varepsilon_0) \leq v(\varepsilon) = v$. In particular, we have (4.10) also in this case. In view of (4.7), this completes the proof. ∎

5 Covering by balls of Lorentz spaces

So far we imposed no assumptions on the set $A \subset \mathbb{R}^n$ which we covered. If A happens to be bounded in some norm $\|\cdot\|$, a new phenomenon occurs. The covering numbers of A by balls in any norm slightly weaker than $\|\cdot\|$ become independent of the dimension n; the parameter that essentially controls them is the combinatorial dimension of A.

This phenomenon is best expressed in the functional setting for Lorentz norms (2.8), because they are especially easy to compare. Given two generating functions ϕ and ψ, we look at their *comparison function*

$$(\phi|\psi)(t) = \sup\{\phi(s) \mid \phi(s) \geq \psi(ts)\}.$$

Fix a probability space (Ω, μ). The comparison function helps us measure to what extent the norm in $\Lambda_\phi = \Lambda_\phi(\Omega, \mu)$ is weaker than the norm in $\Lambda_\psi = \Lambda_\psi(\Omega, \mu)$.

Just for the normalization, we assume that

$$\phi(1) = \psi(1) = 1. \tag{5.1}$$

Let $2 \leq \alpha < \infty$. We rule out the extremal case by assuming that

$$\phi(s) \leq e^{\alpha^t - \alpha} \quad \text{for } t \geq 1. \tag{5.2}$$

Theorem 5.1 *Let ϕ and ψ be generating functions satisfying* (5.1) *and* (5.2)*. Let F be a class of functions* 1*-bounded in Λ_ψ. Then for $0 < t < 1/2$*

$$D(F, \Lambda_\phi, t) \le C\alpha \; v(F, ct) \cdot \log(\phi|\psi)(t/2)$$

Remarks. 1. No nontrivial estimate is possible when $\phi = \psi$. Indeed, even in the simplest case when Ω is finite and μ is uniform, let us take F to be the collection of the functions $f_\omega = \delta_\omega / \|\delta_\omega\|_{\Lambda_\phi}$, $\omega \in \Omega$, where δ_ω is the function that takes value 1 at ω and 0 elsewhere. Clearly, F is 1-bounded in Λ_ϕ and has combinatorial dimension $d(F, t) = 1$ for any $0 < t < 1$. However, $\|f_\omega - f_{\omega'}\|_{\Lambda_\phi} \ge 1$ for $\omega \ne \omega'$. Hence $D(F, \Lambda_\phi, 1/2) = \log |F| = \log |\Omega|$. This can be arbitrarily large.

2. To see the sharpness of Theorem 5.1, notice that for *some* probability measure μ on Ω,

$$D(F, \Lambda_\phi, t) \ge c \; v(F, Ct).$$

A simple argument can be found in [T 02] Proposition 1.4.

In the extremal case of Theorem 5.1, when F is 1-bounded in L_∞, the comparison function becomes just $\phi(t)$, which gives

Corollary 5.2 *Let ϕ be a generating function satisfying* (5.2) *and such that $\phi(1) = 1$. Let F be a class of functions* 1*-bounded in L_∞. Then for $0 < t < 1/2$*

$$D(F, \Lambda_\phi, t) \le C\alpha \; v(F, ct) \cdot \log \phi(t/2).$$

We use Theorem 5.1 for classical Lorentz spaces $L_{p,\infty} = L_{p,\infty}(\Omega, \mu)$ generated by $\phi(t) = t^p$.

Corollary 5.3 *Let $1 \le p < q \le \infty$. Let F be a class of functions* 1*-bounded in $L_{q,\infty}$. Then for $0 < t < 1/2$*

$$D(F, L_{p,\infty}, t) \le C_{p,q} \; v(F, ct) \cdot \log(1/t)$$

where

$$C_{p,q} = C \left(\frac{p^2 q}{q - p} \right).$$

Proof. We apply Theorem 5.1 to the functions $\phi(t) = t^p$ and $\psi(t) = t^q$. In this case the comparison function becomes $(\phi|\psi)(t) = t^{pq/(p-q)}$. To complete the proof, notice that (5.2) holds with $\alpha = p$. ■

Our main interest is in the L_p spaces, for which we obtain

Corollary 5.4 *Let* $1 \le p < q \le \infty$. *Let* F *be a class of functions* 1*-bounded in* L_q. *Then for* $0 < t < 1/2$

$$D(F, L_p, t) \le C_{p,q}\ v(F, c_{p,q} t) \cdot \log(1/c_{p,q} t) \tag{5.3}$$

where

$$C_{p,q} = C\left(\frac{p^2 q}{q-p}\right), \qquad c_{p,q} = c \min\left(1, \left(\frac{q-p}{p}\right)^{1/p}\right).$$

In the next section, this estimate will be applied in an important partial case, when p is a nontrivial proportion of q. In that case, say if $p \le 0.99q$, inequality (5.3) reads

$$D(F, L_p, t) \le Cp^2\ v(F, ct) \cdot \log(1/t). \tag{5.4}$$

The history of estimates obtained prior to Corollary 5.4 and (5.4) is outlined in Section 2 after (2.10).

Proof. Since F is 1-bounded in L_q, it is also 1-bounded in $L_{q,\infty}$. Let p' be so that $p < p' < q$. Fix an $f \in L_{p',\infty}$ with $\|f\|_{p',\infty} \le 1$. Then

$$\begin{aligned} \|f\|_p^p &\le \int_{\{\omega:\ |f(\omega)| \le 1\}} |f(\omega)|^p d\mu + \int_1^\infty p t^{p-1} \mu\{\omega :\ |f(\omega)| \ge t\} dt \\ &\le 1 + p \int_1^\infty t^{p-1-p'} dt \le \frac{p'}{p'-p}. \end{aligned}$$

Taking the p-th root we conclude that if f, g are t-separated in L_p, then they are $(b_{p,p'} t)$-separated in $L_{p',\infty}$, where

$$b_{p,p'} = \left(\frac{p'-p}{p'}\right)^{1/p}.$$

Thus $D(F, L_p, t) \le D(F, L_{p',\infty}, b_{p,p'} t)$. Then the application of Corollary 5.3 with p' and q gives

$$D(F, L_p, t) \le B_{p',q}\ v(F, b_{p,p'} t) \cdot \log(1/b_{p,p'} t)$$

with

$$B_{p',q} = C\left(\frac{p'^2 q}{q-p}\right).$$

If we choose $p' = \min(2p, \frac{p+q}{2})$ then a direct check shows that

$$B_{p',q} \le C\left(\frac{12p^2 q}{q-p}\right), \quad b_{p,p'} \ge \min\left(\frac{1}{2}, \left(\frac{q-p}{4p}\right)^{1/p}\right).$$

This completes the proof. ■

To prove Theorem 5.1, we first reduce the size of the domain Ω (which can be assumed finite) by means of a probabilistic selection and then apply the covering Theorem 3.1.

In the probabilistic selection, we use a standard independent model. Given a finite set I and a parameter $0 < \delta < 1$, we consider selectors δ_i, $i \in I$, which are independent $\{0,1\}$-valued random variables with $\mathbb{E}\delta_i = \delta$. Then the set $J = \{i \in I : \delta_i = 1\}$ is a random subset of I and its average cardinality is $s = \delta|I|$. We call J a random set of expected cardinality s.

Lemma 5.5 *Let $0 < \varepsilon \le 1$. For $t \ge \varepsilon\delta m$,*

$$\mathbb{P}\Big\{\Big|\sum_{i=1}^{m}(\delta_i - \delta)\Big| > t\Big\} \le 2\exp(-c\varepsilon t)$$

where $c > 0$ is an absolute constant.

Proof. This follows from Prokhorov-Bennett inequality. Let (X_i) be a finite sequence of real valued independent mean zero random variables such that $\|X_i\|_\infty \le a$ for every i. If $b^2 = \sum_i \mathbb{E}X_i^2$, then for all $t > 0$

$$p := \mathbb{P}\Big\{\sum_i X_i > t\Big\} \le \exp\Big[t/a - (t/a + b^2/a^2)\log(1 + at/b^2)\Big] \tag{5.5}$$

which is less than $\exp(-t^2/4b^2)$ if $t \le b^2/2a$ (see e.g. [LT] 6.3).

We apply Prokhorov-Bennett inequality for $X_i = \delta_i - \delta$ and with $a = 1$, $b^2 = \delta m$. Consider two cases:

1) $\varepsilon\delta m \le t \le 8\delta m$. Since in that case $t/16 \le b^2/2a$, we have

$$p \le \mathbb{P}\Big\{\sum_i X_i > t/16\Big\} \le \exp(-t^2/64\delta m) \le \exp(-\varepsilon t/64)$$

because $t \ge \varepsilon\delta m$.

2) $t > 8\delta m$. Then $\log(1 + at/b^2) = \log(1 + t/\delta m) > 2$, hence

$$p \le \exp\Big[-(t/a)\Big(\log(1 + at/b^2) - 1\Big)\Big] < \exp(-t).$$

Thus for all $t > \varepsilon\delta m$ we have $p \le \exp(-c\varepsilon t)$. Repeating the argument for $-X_i$, we conclude the proof. ■

Lemma 5.6 *There exist absolute constants $C, c > 0$ for which the following holds. Let $\gamma > 0$ and let $\mathcal{Q}$ be a system of subsets of $\{1, \dots, n\}$ such that*

$$|S| \ge \gamma n \quad \textit{for all } S \in \mathcal{Q}.$$

If σ is a random subset of $\{1,\dots,n\}$ of expected cardinality k satisfying $|\mathcal{Q}| \le 0.001 \cdot \exp(c\gamma k)$, then with probability at least 0.99 *we have*

$$\frac{|S \cap \sigma|}{|\sigma|} \ge 0.99\,\frac{|S|}{n} \quad \textit{for all } S \in \mathcal{Q}.$$

Proof. Let $0 < \delta < 1/2$ and set $\delta_1, \dots, \delta_n$ to be $\{0,1\}$-valued independent random variables with $\mathbb{E}\delta_i = \delta$ for all i. Let $\delta = k/n$; consider the random set $\sigma = \{i : \ \delta_i = 1\}$. For any set $S \subset \{1,\dots,n\}$, $|S \cap \sigma| = \sum_{i \in S} \delta_i$. By Lemma 5.5 applied to a sum over S instead of $\{1,\dots,m\}$ and with $t = 0.001\delta|\mathcal{Q}|$, there is an absolute constant $c_0 > 0$ such that

$$\mathbb{P}\{|S \cap \sigma| < 0.999\delta|S|\} \le 2\exp(-c_0\delta|S|).$$

Since for every $S \in \mathcal{Q}$, $|S| \ge \gamma n$, then

$$\mathbb{P}\Big\{|S \cap \sigma| < 0.999\frac{k}{n} \cdot |S|\Big\} \le 2\exp(-c_0\gamma k).$$

Therefore

$$\mathbb{P}\Big\{\forall S \in \mathcal{Q},\ |S \cap \sigma| \ge 0.999\frac{k}{n} \cdot |S|\Big\} \ge 1 - 2|\mathcal{Q}|\exp(-c_0\gamma k) \ge 0.998$$

provided $c \le c_0/2$. Also, $|\sigma| \le 1.001k$ with probability at least 0.999 since k can be assumed sufficiently large. This completes the proof. ■

Given a finite set I, we will now work with Lorentz spaces $\Lambda_\phi(I) = \Lambda_\phi(I,\mu)$, where μ is the uniform measure on I. The following two lemmas reduce the size of I while keeping both the boundedness of the class F in the Λ_ψ-norm and the separation of F in the Λ_ϕ-norm.

Lemma 5.7 *Let ψ be a generating function. Let f be a function on a finite set I such that*

$$\|f\|_{\Lambda_\psi(I)} \le 1.$$

If σ is a random subset of I of expected cardinality $k > C$, then with probability at least 0.9 *we have*

$$\|f\|_{\Lambda_\psi(\sigma)} \le C.$$

Proof. Let $a > 2$ be a parameter to be chosen later and let δ, δ_j be as in the proof of Lemma 5.6. For $s \in \mathbb{Z}$ define the set

$$I_s = \{j \in I \ : \ |f(j)| > 2^s\}.$$

Since $\|f\|_{\Lambda_\psi(I)} \le 1$, we have

$$|I_s| \le \frac{n}{\psi(2^s)}. \tag{5.6}$$

Define also the event A_s as

$$A_s = \Big\{ |I_s \cap \sigma| > \frac{a\delta n}{\psi(2^s)} \Big\}.$$

We want to bound the probability that at least one A_s occurs. Let r be the maximal number such that $\delta |I_r| \ge 0.01$. Then

$$\mathbb{P}\{\forall j \in I_{r+1}\ \delta_j = 0\} = (1-\delta)^{|I_{r+1}|} \ge e^{-\delta |I_{r+1}|} \ge e^{-0.01} > 0.99.$$

If A_s occurs for some $s > r$ then $I_s \cap \sigma$ is nonempty, hence the larger set $I_{r+1} \cap \sigma$ is nonempty, which happens with probability at most 0.01. Thus

$$\mathbb{P}\Big(\bigcup_{s>r} A_s \Big) \le 0.01. \tag{5.7}$$

To bound $\mathbb{P}(A_s)$ with $s \le r$, we will apply Lemma 5.5 with $m = |I_s|$ and $t = \frac{a\delta n}{2\psi(2^s)}$. Note that

$$\mathbb{P}(A_s) = \mathbb{P}\Big\{ \sum_{j \in I_s} (\delta_j - \delta) > \frac{a\delta n}{\psi(2^s)} - \delta m \Big\}.$$

By (5.6), we have $\frac{a\delta n}{\psi(2^s)} - \delta m \ge t \ge \delta m$. Then Lemma 5.5 gives

$$\mathbb{P}(A_s) \le \exp\Big(- \frac{ca\delta n}{\psi(2^s)} \Big).$$

By the convexity of ψ,

$$\psi(wx) \ge w\psi(x) \quad \text{for all } x \ge 0 \text{ and } w \ge 1. \tag{5.8}$$

Thus $\psi(2^s) \le 2^{s-r}\psi(2^r)$. Then using (5.6) and the fact that $\delta|I_r| \ge 0.01$, we obtain

$$\mathbb{P}(A_s) \le \exp\Big(- 2^{r-s} \frac{ca\delta n}{\psi(2^r)} \Big) \le \exp(-2^{r-s} ca\delta |I_r|) \le \exp(-0.01 ca 2^{r-s}).$$

So if a is taken large enough then $\sum_{s=-\infty}^{r} \mathbb{P}(A_s) \le 0.01$. Combining this with (5.7), we conclude that

$$\mathbb{P}\Big(\bigcup_{s \in \mathbb{Z}} A_s \Big) \le 0.02.$$

In addition, by Lemma 5.5 we have $\mathbb{P}\{|\sigma| < \frac{1}{2}\delta n\} \le 0.02$ since k can be assumed large enough.

Now suppose that $|\sigma| \ge \frac{1}{2}\delta n$ and that none of the events A_s occur, which happens with probability at least $1 - 0.02 - 0.02 = 0.96$. Fix any $t > 0$ and find an integer s so that $2^s \le t < 2^{s+1}$. By the definitions of A_s, I_s and by (5.8),

$$|\{i \in \sigma : |f(i)| > t\}| \le |I_s \cap \sigma| \le \frac{a\delta n}{\psi(2^s)} \le \frac{2a|\sigma|}{\psi(2^s)} \le \frac{2a|\sigma|}{\psi(t/2)} \le \frac{|\sigma|}{\psi(t/4a)}.$$

This means that $\|f\|_{\Lambda_\psi(\sigma)} \le 4a$. ∎

Lemma 5.8 *Let ϕ, ψ be Lorentz functions. Let F be a class of functions on a finite set I, which is* 1*-bounded in the $\Lambda_\psi(I)$ norm. Assume that*

$$\|x\|_{\Lambda_\phi(I)} \ge t \quad \text{for all } x \in F. \tag{5.9}$$

If σ is a random subset of I of expected cardinality k satisfying $|F| \le 0.001 \exp\left(\frac{ck}{(\phi|\psi)(t)}\right)$, then with probability at least 0.99 *we have*

$$\|x\|_{\Lambda_\phi(\sigma)} \ge 0.99\|x\|_{\Lambda_\phi(I)} \quad \text{for all } x \in F.$$

This lemma will be applied to the difference set $A - A$ of a t-net A of the class F in the theorem.

Proof. We can assume that $I = \{1, \dots, n\}$. Fix an $x \in F$. Since $|x|/\|x\|_{\Lambda_\phi(I)} = 1$, there exists an $s = s(x) > 0$ such that

$$\mu\Big\{\frac{|x|}{\|x\|_{\Lambda_\phi(I)}} > s\Big\} \ge \frac{1}{\phi(s)}. \tag{5.10}$$

On the other hand, since $\|x\|_{\Lambda_\phi(I)} \ge t$ and $\|x\|_{\Lambda_\psi(I)} \le 1$, the measure in (5.10) is majorized by

$$\mu\{|x| > ts\} \le \frac{1}{\psi(ts)}.$$

Hence $\phi(s) \ge \psi(ts)$ and therefore

$$\phi(s) \le (\phi|\psi)(t). \tag{5.11}$$

Now consider the family of subsets of I defined as

$$S(x) = \Big\{i : \frac{|x(i)|}{\|x\|_{\Lambda_\phi(I)}} > s(x)\Big\}, \quad x \in F.$$

By (5.10) and (5.11), for every $x \in F$

$$|S(x)| \geq \frac{n}{\phi(s(x))} \geq \frac{n}{(\phi|\psi)(t)}.$$

Let μ_σ denote the uniform probability measure on σ. Lemma 5.6 implies that whenever $|F| \leq 0.001 \exp\big(\frac{ck}{(\phi|\psi)(t)}\big)$, a random subset σ of I of average cardinality k satisfies with probability at least 0.99 that

$$\begin{aligned}\mu_\sigma\Big\{\frac{|x|}{\|x\|_{\Lambda_\phi(I)}} > s(x)\Big\} &= \frac{|S(x) \cap \sigma|}{|\sigma|} \geq 0.99\frac{|S(x)|}{n} \\ &\geq 0.99\frac{1}{\phi(s(x))} \geq \frac{1}{\phi(s(x)/0.99)} \quad \text{for all } x \in F.\end{aligned}$$

Hence

$$\|x\|_{\Lambda_\phi(\sigma)} \geq 0.99\|x\|_{\Lambda_\phi(I)} \quad \text{for all } x \in F.$$

The proof is complete. ■

Proof of Theorem 5.1. We may assume that Ω is finite. By splitting the atoms of Ω (by replacing an atom ω by, say, two atoms ω_1 and ω_2, each carrying measure $\frac{1}{2}\mu(\omega)$ and by defining $f(\omega_1) = f(\omega_2) = f(\omega)$ for $f \in F$), we can make the measure μ almost uniform without changing neither the covering numbers nor the combinatorial dimension of F. So, we can assume that μ is the uniform measure on Ω.

Let A be a t-separated subset of F (which means that $\|f - g\|_{\Lambda_\phi(\Omega)} \geq t$ for all $f \neq g$ in F) of size

$$\log|A| = D(F, \Lambda_\phi(\Omega), t).$$

The difference set $\frac{1}{2}(A - A) \setminus \{0\} = \{\frac{1}{2}(f - g) : f \neq g;\ f, g \in A\}$ satisfies the assumptions of Lemma 5.8 with $t/2$ in (5.9). Then for k defined by

$$|A|^2 = 0.001 \exp\Big(\frac{ck}{(\phi|\psi)(t/2)}\Big), \tag{5.12}$$

a random subset σ of Ω of average cardinality k satisfies with probability at least 0.99 that

$$\Big\|\frac{1}{2}(f - g)\Big\|_{\Lambda_\phi(\sigma)} \geq 0.99\Big\|\frac{1}{2}(f - g)\Big\|_{\Lambda_\phi(\Omega)} \geq 0.99\frac{t}{2} \geq \frac{t}{3} \quad \text{for all } f \neq g \text{ in } A.$$

This means that

$$A \text{ is } \frac{t}{3}\text{-separated in } \Lambda_\phi(\sigma) \tag{5.13}$$

and in particular

$$D(F, \Lambda_\phi(\sigma), t/3) \geq \log|A| = D(F, \Lambda_\phi(\Omega), t).$$

The advantage of the left hand side is that the size of σ is controlled via (5.12).

We need also to keep A well bounded in $\Lambda_\psi(\sigma)$. Denote by $\mathbb{E}_\sigma$ the average over the random set σ, that is over the selectors δ_i. By Lemma 5.7

$$\mathbb{E}\,|\{f \in A\ : \|f\|_{\Lambda_\psi(\sigma)} \leq C\}| = \sum_{f \in A} \mathbb{P}(\|f\|_{\Lambda_\psi(\sigma)} \leq C) \geq 0.9|A|.$$

Therefore with probability at least 0.8,

$$\text{at least a half of the functions in } A \text{ have norm } \|f\|_{L_\psi(\sigma)} \leq C. \tag{5.14}$$

Since $k/2 \leq |\sigma| \leq 2k$ holds with probability at least 0.9, there exists a realization of σ that satisfies simultaneously this property, (5.13) and (5.14). Let B be the set consisting of $\frac{6}{t}f$, where f are the functions satisfying (5.14).

Summarizing, there exist a subset σ of Ω and a set B such that

- B is a subset of $\frac{6}{t}A$,
- B is (C/t)-bounded in $L_\psi(\sigma)$,
- B is 2-separated in $\Lambda_\phi(\sigma)$,
- $|B| \geq |A|/2 \geq c' \exp\left(\frac{c|\sigma|}{2(\phi|\psi)(t/2)}\right)$.

We can clearly assume that $\sigma = \{1, \ldots, n\}$ and realize the space $\Lambda_\phi^n = \Lambda_\phi(\{1, \ldots, n\})$ as $\mathbb{R}^n$ equipped with the Lorentz norm Λ_ϕ. Applying the covering Theorem 3.1, we get

$$N(B, \mathrm{Tower}^\alpha) \leq \Sigma(B)^\alpha. \tag{5.15}$$

Since B is 2-separated in Λ_ϕ^n and by (5.2) the norm in this space is bounded by the Tower^α norm, the set B is also 2-separated in the Tower^α norm. Hence

$$N(B, \mathrm{Tower}^\alpha) = |B|. \tag{5.16}$$

The right hand side of (5.15) can be estimated through Lemma 4.7. By (5.1) and convexity, $\psi(t) \geq t$ for $t \geq 1$. Thus $C\|f\|_{L_\psi} \geq \|f\|_{L_{1/2}}$ for all functions f. Hence

$$B \subseteq (C/t)\,\mathrm{Ball}(L_\psi^n) \subseteq (C'/t)\,\mathrm{Ball}(L_{1/2}^n).$$

Hence by (5.15), (5.16) and Lemma 4.7,

$$|B| \le \Big(\frac{Cn}{tv}\Big)^{3v}, \quad \text{where } v = v(B) = v(B,1) \le v(A, t/6). \tag{5.17}$$

We also have a lower bound $|B| \ge c' \exp(an)$ with $a = \frac{c/2}{(\phi|\psi)(t/2)}$. Taking logarithms of the upper and the lower bounds, we obtain $an/v \le C \log(Cn/tv)$, from which it follows that

$$\frac{n}{v} \le \frac{C}{a} \log\Big(\frac{C}{ta}\Big).$$

Plugging this back into (5.17), we obtain

$$|B| \le \Big[\Big(\frac{C}{ta}\Big) \log\Big(\frac{C}{ta}\Big)\Big]^{Cv} \le \Big(\frac{C}{ta}\Big)^{C_1 v}.$$

Note that

$$(\phi|\psi)(t) \ge 1/t \quad \text{for } 0 < t < 1.$$

Indeed, $\phi(\frac{1}{t}) \ge 1 = \psi(t \cdot \frac{1}{t})$ which implies $(\phi|\psi)(t) \ge \phi(\frac{1}{t}) \ge \frac{1}{t}$.

Therefore $t \ge (2/c)a$ and finally

$$|B| \le a^{-Cv} \le (\phi|\psi)(t/2)^{Cv}.$$

On the other hand, by the construction

$$\log|B| \ge \log\Big(\frac{1}{2}|A|\Big) \ge cD(F, \Lambda_\phi(\Omega), t).$$

This completes the proof. ■

6 Random processes and the uniform entropy

Here we prove our main results, Theorems 1.2 and 1.3, which compare the uniform entropy $D(F,t)$ to the combinatorial dimension $v(F,t)$. One direction of this comparison is easy: for every class of functions F and every $t > 0$,

$$D(F,t) \ge c\, v(F, 2t) \tag{6.1}$$

where $c > 0$ is an absolute constant, see [T 02].

The reverse inequality is not true in general even for $\{0,1\}$ classes. Let, for example, F be the collection of n characteristic functions $\mathbf{1}_{\{i\}}$ of the singletons $i \in \{1, \dots, n\}$. Then for $0 < t < n^{-1/2}$ we have $D(F,t) = \log n$ while $v(F,t) = 1$.

Nevertheless, we are able to show that the reverse to (6.1) holds: 1) under a minimal regularity of F, and 2) always after taking integrals on both sides.

Integral equivalence The following is a general form of Theorem 1.2.

Theorem 6.1 *For every class F and for any $b \geq 0$,*

$$\int_b^\infty \sqrt{D(F,t)}\,dt \leq C \int_{cb}^\infty \sqrt{v(F,t)}\,dt. \tag{6.2}$$

The proof of Theorem 6.1 is based on the following

Lemma 6.2 *Let $a > 2$ and let F be a function class. Then for all $t > 0$:*

$$D(F,t) \leq C \log a \sum_{j=0}^{\infty} 4^j\, v(F, ca^j t). \tag{6.3}$$

Proof. The proof of the Lemma uses an iteration argument. It relies on the fact valid for arbitrary sets K, D and L in $\mathbb{R}^n$:

$$N(K,D) \leq N(K,L) \sup_{z \in \mathbb{R}^n} N((K+z) \cap L, D). \tag{6.4}$$

To check this, first cover K by translates of L and then cover the intersection of K with each translate by appropriate translates of D.

It will be easier to work with the "covering" analog of $D(F,t)$, so we define a covering version of $D(F,X,t)$ in (2.5) as

$$D'(F,X,t) = \log \sup \Big(n \mid \exists f_1, \ldots, f_n \in X \ \forall f \in F \ \exists i \ \|f - f_i\|_X \leq t \Big).$$

By (2.7),

$$D(F,X,2t) \leq D'(F,X,t) \leq D(F,X,t). \tag{6.5}$$

We can clearly assume the domain Ω to be finite. Fix the underlying probability μ on Ω and $t > 0$. For $j = 1, 2, \ldots$ define

$$t_j = a^{j-1} t \quad \text{and} \quad X_j = L_{2^j}(\Omega, \mu).$$

We estimate $D(F, L_2(\Omega,\mu), t)$ by (6.4):

$$D'(F, X_1, t_1) \leq D'(F, X_2, t_2) + \sup_h D'\big((F+h) \cap t_2 \mathrm{Ball}(X_2), X_1, t_1\big)$$

where the supremum is over all functions h on the (finite) domain Ω. Iterating this inequality, we obtain

$$D'(F, X_1, t_1) \leq \sum_{j=1}^{\infty} \sup_h D'(F_j(h), X_j, t_j) \tag{6.6}$$

where $F_j(h) = (F+h) \cap t_{j+1}\mathrm{Ball}(X_{j+1})$. Obviously, the class $F_j(h)$ is t_{j+1}-bounded in X_{j+1}. Then applying (5.4) to the class $t_{j+1}^{-1} F_j(h)$ with $p = 2^j$ and $q = 2^{j+1}$, we obtain

$$D'(F_j(h), X_j, t_j) \le C 4^j v(F_j(h), ct_j) \cdot \log(t_{j+1}/t_j) \le C 4^j v(F, ct_j) \cdot \log a$$

because apparently $v(F_j(h), s) \le v(F, s)$ for every s. To complete the proof we substitute the previous inequality into (6.6) and use (6.5). ■

Proof of Theorem 6.1. Applying Lemma 6.2 and Jensen's inequality, we have for $a = 3$:

$$\int_b^\infty \sqrt{D(F,t)}\, dt \le C\sqrt{\log a} \sum_{j=0}^\infty 2^j \int_b^\infty \sqrt{v(F, ca^j t)}\, dt$$

$$= C\sqrt{\log a} \sum_{j=0}^\infty (2/a)^j \int_{ca^j b}^\infty \sqrt{v(F,u)}\, du \le C \int_{cb}^\infty \sqrt{v(F,u)}\, du.$$

■

Remark. The square roots in (6.2) can of course be replaced by any other equal positive powers. The present form of (6.2) was chosen to match Dudley's entropy integral, see Theorem 6.5 below.

Theorem 1.1 follows from Theorem 6.1 as explained in the introduction. The gap between the sufficient and the necessary conditions in Theorem 1.1 is known to be needed in general (at least in the description of universal Donsker classes, see [Du 99] Propositions 10.1.8 and 10.1.14).

Pointwise equivalence A similar argument, which we give now, completes the proof of the other main result of the paper, Theorem 1.3. Although (6.1) can not be reversed in general, a remarkable fact is that it can be reversed if the combinatorial dimension is polynomial in t.

Theorem 6.3 *Let F be a class of functions and $t > 0$. Assume that there exist positive numbers v and $\alpha \le 1$ such that*

$$v(F, tx) \le v x^{-\alpha} \quad \text{for all } x \ge 1. \tag{6.7}$$

Then

$$D(F, Ct) \le (C/\alpha)\, v.$$

Proof. Applying Lemma 6.2 with $a = 5^{1/\alpha}$ and estimating the combinatorial dimension via (6.7), we have

$$v(F, t/c) \le C \log a \sum_{j=0}^{\infty} 4^j v a^{-\alpha j} \le (C/\alpha)\, v.$$

This completes the proof. ■

The following is a general form of Theorem 1.3. It improves upon Talagrand's inequality proved in [T 87], see [T 02].

Corollary 6.4 *Let F be a class of functions and $t > 0$. Assume that there exists a decreasing function $v(t)$ and a number $a > 2$ such that*

$$v(F, s) \le v(s) \quad \text{and} \quad v(as) \le \frac{1}{2}\, v(s) \quad \text{for all } s \ge t. \tag{6.8}$$

Then

$$D(F, Ct) \le C \log a \cdot v(t).$$

Proof. Applying (6.8) recursively, we have $v(a^j t) \le \frac{1}{2^j}\, v(t)$ for all $j = 0, 1, 2, \ldots$. Let $x \ge 1$ and choose j so that $a^j \le x \le a^{j+1}$. Then

$$2^{-j} \ge x^{-\frac{\log 2}{\log a}} \ge 2^{-j-1},$$

so,

$$v(F, tx) \le v(tx) \le v(ta^j) \le C v(t) \cdot 2^{-j} \le 2C\, v(t) \cdot x^{-\frac{\log 2}{\log a}}.$$

The conclusion follows by Theorem 6.3. ■

A combinatorial bound on Gaussian processes A quantitative version of Theorem 1.1 is the following bound on Gaussian processes indexed by F in terms of the combinatorial dimension of F.

Let F be a class of functions on an n-point set I. The standard Gaussian process indexed by $f \in F$ is

$$X_f = \sum_{i \in I} g_i f(i)$$

where (g_i) are independent $N(0, 1)$ random variables. The problem is to bound the supremum of the process (X_f) normalized by the standard deviation as in the Central Limit Theorem:

$$E(F) = n^{-1/2}\, \mathbb{E} \sup_{f \in F} X_f.$$

Theorem 6.5 *For every class F,*

$$E(F) \leq C \int_0^\infty \sqrt{v(F,t)}\, dt \tag{6.9}$$

where C is an absolute constant. Moreover, 0 can be replaced by $cn^{-1/2}E(F)$, where $c > 0$ is an absolute constant.

Proof. By Dudley's entropy integral inequality,

$$E(F) \leq C \int_{n^{-1/2}E(F)}^\infty \sqrt{D(F, L_2(\mu), t)}\, dt \tag{6.10}$$

where μ is the uniform probability measure on I, see [MV 03]. Then the proof is completed by Theorem 6.1. ■

In 1992, M. Talagrand proved Theorem 6.5 for uniformly bounded convex classes and up to an additional factor of $\log^M(1/t)$ in the integrand; this was a main result of [T 92]. The absolute constant M was reduced to 1/2 in [MV 03]. Theorem 6.5 is optimal. We emphasize its important meaning:

> *In the classical Dudley's entropy integral, the entropy can be replaced by the combinatorial dimension.*

Optimality of the bound on Gaussian processes We conclude this section by showing the sharpness of Theorem 6.5. For every n one easily finds a class F for which the inequality in (6.9) can be reversed – this is true e.g. for $F = \{-1, 1\}^I$. More importantly, the integral in (6.9) can not be improved in general to the (Sudakov-type) supremum $\sup t\sqrt{v(F,t)}$. This is so even if we replace the Gaussian process X_f by the Rademacher process

$$Y_f = \sum_{i \in I} \varepsilon_i f(i)$$

where (ε_i) are independent symmetric ± 1 valued random variables. The average supremum of such process,

$$E_{rad}(F) = n^{-1/2}\, \mathbb{E} \sup_{f \in F} Y_f,$$

is well known to be majorized by that of the Gaussian process: $E_{rad}(F) \leq CE(F)$ (see [LT] Lemma 4.5).

Proposition 6.6 *For every n, there exists a class F of functions on $\{1,\dots,n\}$ uniformly bounded by 1 and such that*

$$E_{rad}(F) \geq c_1 \int_0^\infty \sqrt{v(F,t)}\,dt \geq c \log n \cdot \sup_{t>0} t\sqrt{v(F,t)}. \tag{6.11}$$

Our example will be constructed as sums of random vertices of the discrete cube with quickly decreasing weights.

We shall bound $E_{rad}(F)$ from below via a Sudakov type minoration for Rademacher processes. Let $D = \mathrm{Ball}(L_2^n) = \sqrt{n}B_2^n$. Proposition 4.13 of [LT] with $\varepsilon = \frac{1}{2}n$ states the following:

Fact 6.7 *If A is a subset of $\mathbb{R}^n$ and*

$$\sup_{x\in A} \|x\|_\infty \leq c_1 \frac{\sqrt{n}}{E_{rad}(A)}, \tag{6.12}$$

then

$$\sqrt{\log N\big(A, \frac{1}{2}D\big)} \leq C E_{rad}(A). \tag{6.13}$$

The entropy in (6.13) will be estimated in a standard way:

Fact 6.8 *There exists an absolute constant α such that the following holds. Let A be a set of $N \leq e^{\alpha n}$ random vertices of the discrete cube $\{-1,1\}^n$, i.e. A consists of N independent copies of a random vector $(\varepsilon_1,\dots,\varepsilon_n)$. Then with probability at least 1/2,*

$$N(A, \frac{1}{2}D) \geq \sqrt{N}.$$

Proof. Obviously, we can assume that $N \geq 2$. Assume that the event

$$N(A, \frac{1}{2}D) \leq \sqrt{N} \tag{6.14}$$

occurs. Then there exists a translate $D'_x = \frac{1}{2}D + x$ of $\frac{1}{2}D$ which contains at least $N/N(A,\frac{1}{2}D) \geq \sqrt{N}$ points from A. Set $A' = A \cap D'_x$. By dividing the set A' into pairs in an arbitrary way, we can find a set $\mathcal{P}$ of $M \geq \frac{\sqrt{N}-1}{2}$ pairs $(x,y) \in A' \times A'$, $x \neq y$, so that each point from A belongs to at most one pair in $\mathcal{P}$. Since A' lies in a single translate of $\frac{1}{2}D$, we have

$$\|x-y\|_{L_2^n} \leq 1 \quad \text{for all } (x,y) \in \mathcal{P}. \tag{6.15}$$

Thus if (6.14) occurs, then (6.15) occurs for some M-element set $\mathcal{P} \subset A \times A$.

Let now $\mathcal{P}$ be a *fixed* set of M disjoint pairs of elements of A. Then

$$\mathbb{P}(\text{event (6.15) occurs}) = (\mathbb{P}(\|x-y\|_{L_2^n} \le 1))^M \tag{6.16}$$

where x and y are independent random vertices of the discrete cube. Here we used the fact that the pairs in $\mathcal{P}$ are disjoint from each other and, consequently, are jointly independent. The probability in (6.16) is easily estimated using Prokhorov-Bennett inequality (5.5):

$$\mathbb{P}(\|x-y\|_{L_2^n} \le 1) = \mathbb{P}\Big(\sum_{i=1}^n |\varepsilon_{1i} - \varepsilon_{2i}|^2 \le n\Big) \le e^{-c_1 n}$$

where (ε_{1i}) and (ε_{2i}) are independent copies of the random vector (ε_i).

To estimate the probability that the event (6.14) occurs, note that there is less than $\binom{N^2}{M}$ ways to choose $\mathcal{P}$. Therefore

$$p := \mathbb{P}\Big(N(A, \frac{1}{2}D) \le \sqrt{N}\Big) \le \binom{N^2}{M}(e^{-c_1 n})^M \le \Big(\frac{eN^2}{M}e^{-c_1 n}\Big)^M \le \big(Ne^{-c_2 n}\big)^{3M/2}.$$

Since $N \ge 2$, we have $M \ge 1$. If $\alpha \le c_2/2$, we conclude that $p \le \exp(-\frac{c_2 n}{2} \cdot 3M/2) < 1/2$. This completes the proof. ■

Corollary 6.9 *There exists an absolute constant α such that the following holds. Let A be a set of $N \le e^{\alpha n}$ random vertices of the discrete cube. Then with probability at least $1/2$,*

$$c\sqrt{\log N} \le E_{rad}(A) \le C\sqrt{\log N}.$$

Proof. Since $A \subset \{-1,1\}^n \subset \sqrt{n}B_2^n$, we have

$$E_{rad}(A) \le CE(A) \le C_1\sqrt{\log N}, \tag{6.17}$$

see [LT] (3.13). To prove the reverse inequality, assume that $N \le e^{\beta n}$ for $\beta = \min(\alpha, (c_1/C_1)^2)$, where c', c_1 and C_1 are the absolute constants from Fact 6.8, (6.12) and (6.17) respectively. Then (6.17) implies that (6.12) is satisfied. Hence by Facts 6.7 and 6.8, with probability at least $1/2$ we have

$$E_{rad}(A) \ge (1/C)\sqrt{\log N(A, \frac{1}{2}D)} \ge (c/\sqrt{2})\sqrt{\log N}.$$

This completes the proof. ■

Proof of Proposition 6.6. Fix a positive integer n. Let k_1 be the maximal integer so that $2^{4^{k_1}} \leq e^{\alpha n}$, where α is an absolute constant from Corollary 6.9. For each $1 \leq k \leq k_1$ define a set F_k in $\mathbb{R}^n$ as follows. Let $N(k) = 2^{4^k}$ and let $A_k = \{x_1^k, \dots, x_{N(k)}^k\}$ be a family of points in the discrete cube $\{-1,1\}^n$ in $\mathbb{R}^n$ satisfying Corollary 6.9. Set

$$F_k = 2^{-k} \cdot A_k \quad \text{and put} \quad F = \sum_{k=1}^{k_1} F_k$$

where the sum is the Minkowski sum: $A + B = \{a + b \,:\, a \in A, b \in B\}$. Then F is a uniformly bounded class of functions on $\{1, \dots, n\}$.

By Corollary 6.9 we have

$$E_r(F) = \sum_{k=1}^{k_1} E_r(F_k) \geq c \sum_{k=1}^{k_1} 2^{-k} \sqrt{\log N(k)} \geq c k_1 \geq c_1 \log n. \tag{6.18}$$

To estimate the combinatorial dimension of F, fix a $t = 2^{-k}$ with $0 \leq k \leq k_1$. Then

$$v(F,t) \leq v\Big(\sum_{l=1}^{k+1} F_l, \frac{t}{2}\Big),$$

because the diameter of $\sum_{l=k+2}^{k_1} F_l$ in the L_∞ norm is at most $t/2$. Obviously, by definition of the combinatorial dimension, for any finite set F we have $v(F,t) \leq \log_2 |F|$, so

$$v(F,t) \leq \log_2 \Big| \sum_{l=1}^{k+1} F_l \Big| = \sum_{l=1}^{k+1} \log_2 |F_l| = \sum_{l=1}^{k+1} 4^l \leq C 4^k. \tag{6.19}$$

This shows that $t\sqrt{v(F,t)} \leq C_1$ for all $t \geq 2^{-k_1}$. Since $2^{-k_1} \leq 2/\sqrt{n}$ and clearly $t\sqrt{v(F,t)} \leq t\sqrt{n} \leq 2$ for all $t \leq 2/\sqrt{n}$, we conclude that

$$\sup_{t>0} t\sqrt{v(F,t)} \leq C_1. \tag{6.20}$$

Also, since F is uniformly bounded by 1, we have $v(F,t) = 0$ for all $t > 1$. Moreover, for all $f, g \in F$ and all $i \in \{1, \dots, n\}$, we have $|f(i) - g(i)| \geq 2^{-k_1}$ whenever $f(i) \neq g(i)$. Hence, $v(F,t) = v(F,t_1)$ for all $t \leq t_1 = 2^{-k_1}$. Thus,

$$\int_0^\infty \sqrt{v(F,t)}\, dt \leq C \sum_{k=0}^{k_1} 2^{-k} \sqrt{v(F, 2^{-k})} \leq C k_1 \leq C \log n.$$

This, (6.18) and Theorem 6.5 imply that the leftmost and the middle quantities in (6.11) are both equivalent to $\log n$ up to an absolute constant factor. Together with (6.20), this completes the proof. ■

7 Sections of convex bodies

First applications of entropy inequalities involving the combinatorial dimension to geometric functional analysis are due to M. Talargand [T 92]. Using his entropy inequality (which (6.9) strengthens) he proved that for Banach spaces infratype p implies type p ($1 < p < 2$). He also proved the classical Elton Theorem with asymptotics that fell short from optimal, improving earlier estimates by J.Elton [E] and A.Pajor [Pa 85]; the optimal asymptotics were found in [MV 03] using (2.10).

Here we will use new covering results to find nice coordinate sections of a general convex body (for simplicity, we will assume that the body is symmetric with respect to the origin). Our main result is related to three classical results in geometric functional analysis – Dvoretzky's Theorem in the form of V. Milman (see [MS] 4.2), Bourgain-Tzafriri's Principle of the Restricted Invertibility [BT 87] and Elton's Theorem ([E], see also [Pa 85], [T 92], [MV 03]).

By B_p^n we denote the unit ball of l_p^n, that is the set of all $x \in \mathbb{R}^n$ such that $\sum_1^n |x(i)|^p \le 1$. Let K be a convex body symmetric with respect to the origin. Its average size is measured by $M_K = \int_{S^{n-1}} \|x\|_K \; d\sigma(x)$, where σ is the normalized Lebesgue measure on the sphere S^{n-1} and $\|x\|_K$ denotes the Minkowski functional of K (the seminorm whose unit ball is K).

Theorem 7.1 (Dvoretzky's Theorem, see [MS]) *Let K be a symmetric convex body in $\mathbb{R}^n$ containing B_2^n. Then there exists a subspace E in $\mathbb{R}^n$ of dimension $k \ge cM_K^2 n$ and such that*

$$c(B_2^n \cap E) \subseteq M_K(K \cap E) \subseteq C(B_2^n \cap E). \tag{7.1}$$

Moreover, a random subspace E taken uniformly in the Grassmanian $G_{n,k}$ satisfies (7.1) *with probability at least $1 - e^{-ck}$.*

Next theorem, the Principle of the Restricted Invertibility due to J. Bourgain and L. Tzafriri, is the first and probably the most used result from the extensive paper [BT 87]. By (e_i) we denote the canonical basis of $\mathbb{R}^n$.

Theorem 7.2 (J. Bourgain and L. Tzafriri [BT 87]) *Let $T : l_2^n \to l_2^n$ be a linear operator with $\|Te_i\| \ge 1$ for all i. Then there exists a subset σ of $\{1, \dots, n\}$ of size $|\sigma| \ge cn/\|T\|^2$ and such that*

$$\|Tx\| \ge c\|x\| \quad \textit{for all } x \in \mathbb{R}^\sigma.$$

Denote by (ε_n) Rademacher random variables, i.e. sequence of independent symmetric ± 1 valued random variables.

Theorem 7.3 (J. Elton [E]) *Let $x_1, \ldots, x_n$ be vectors in a real Banach space, satisfying*

$$\forall i \ \|x_i\| \le 1 \quad \textit{and} \quad \mathbb{E}\Big\| \sum_{i=1}^n \varepsilon_i x_i \Big\| \ge \delta n$$

for some number $\delta > 0$. Then there exists a subset $\sigma \subset \{1, \ldots, n\}$ of cardinality $|\sigma| \ge c_1(\delta) n$ such that

$$\Big\| \sum_{i \in \sigma} a_i x_i \Big\| \ge c_2(\delta) \sum_{i \in \sigma} |a_i|$$

for all real numbers (a_i).

The best possible asymptotics is known: $c_1(\delta) \asymp \delta^2$ and $c_2(\delta) \asymp \delta$ [MV 03].

Now we state our main result. By (g_i) we denote a sequence of independent normalized Gaussian random variables.

Theorem 7.4 *Let $x_1, \ldots, x_n$ be vectors in a real Banach space, satisfying*

$$\Big\| \sum_{i=1}^n a_i x_i \Big\| \le \sqrt{n} \Big(\sum_{i=1}^n |a_i|^2 \Big)^{1/2} \quad \textit{and} \quad \mathbb{E}\Big\| \sum_{i=1}^n g_i x_i \Big\| \ge \delta n \tag{7.2}$$

for all real numbers (a_i) and for some number $\delta > 0$. Then there exist two numbers $s > 0$ and $c\delta \le t \le 1$ connected by the inequality $st \ge c\delta / \log^{3/2}(2/\delta)$ and a subset σ of $\{1, \ldots, n\}$ of size $|\sigma| \ge s^2 n$ such that

$$\Big\| \sum_{i \in \sigma} a_i x_i \Big\| \ge ct \sum_{i \in \sigma} |a_i| \tag{7.3}$$

for all real numbers (a_i).

The first assumption in (7.2) is satisfied in particular if $\|x_i\| \le 1 \ \forall i$. Also, since $t \le 1$, we always have $s \ge c\delta / \log^{3/2}(2/\delta)$. This instantly recovers Elton's Theorem.

Next, Theorem 7.4 essentially extends the Bourgan-Tzafriri principle of restricted invertibility to operators $T : l_2^n \to X$ acting into arbitrary Banach space X. The average size of T is measured by its ℓ-norm defined as $\ell(T)^2 = \mathbb{E}\|Tg\|^2$, where $g = (g_1, \ldots, g_n)$. If X is a Hilbert space, then $\ell(T)$ equals the Hilbert-Schmidt norm of T.

Corollary 7.5 (General Principle of the Restricted Invertibility) *Let $T : l_2^n \to X$ be a linear operator with $\ell(T) \ge \sqrt{n}$, where X is a Banach space. Let $\alpha = c \log^{-3/2}(2\|T\|)$. Then there exists a subset σ of $\{1, \ldots, n\}$ of size $|\sigma| \ge c\alpha^2 n / \|T\|^2$ and such that*

$$\|Tx\| \ge \alpha |\sigma|^{-1/2} \|x\|_{l_1^\sigma} \quad \textit{for all } x \in \mathbb{R}^\sigma.$$

If X is a Hilbert space, the condition $\ell(T) \geq \sqrt{n}$ is satisfied, for example, if $\|x_i\| \geq 1$ for all i. In that case $|\sigma|^{-1/2}\|x\|_{l_1^\sigma}$ in the conclusion can be improved to $|x\| = \|x\|_{l_2^n}$ via the Grothendieck factorization (we will do this below). This recovers the Bourgain-Tzarfiri Theorem up to the logarithmic factor α.

Proof of Corollary 7.5. We apply Theorem 7.4 to the vectors $x_i = \frac{\sqrt{n}}{\|T\|} Te_i$, $i = 1, \dots, n$. Then for any $a_1, \dots a_n$

$$\Big\|\sum_{i=1}^n a_i x_i\Big\| = \frac{\sqrt{n}}{\|T\|} \cdot \Big\|T(\sum_{i=1}^n a_i e_i)\Big\| \leq \sqrt{n} \cdot \|(a_1, \dots a_n)\|_{\ell_2^n}.$$

Since by Kahane's inequality $\ell(T) \leq C\,\mathbb{E}\|Tg\|$, the second assumption in (7.2) holds with $\delta = c/\|T\|$. Hence there exist numbers $c/\|T\| \leq t \leq 1$ and s satisfying

$$st \geq c\delta/\log^{3/2}(2/\delta)$$

and a subset σ of $\{1, \dots, n\}$ of size $|\sigma| \geq s^2 n$ so that we have (multiplying both sides by $|\sigma|^{-1/2}$)

$$\frac{C\sqrt{n/|\sigma|}}{\|T\|}\Big\|\sum_{i\in\sigma} a_i Te_i\Big\| \geq ct|\sigma|^{-1/2}\sum_{i\in\sigma}|a_i| \quad \text{for all real numbers } (a_i).$$

Since $t \leq 1$, we have $s \geq c\delta/\log^{3/2}(2/\delta) \geq \alpha/\|T\|$ and consequently $|\sigma| \geq \alpha^2 n/\|T\|^2$ as required. As $|\sigma| \geq s^2 n$, we have $s \leq \sqrt{|\sigma|/n}$, hence

$$\frac{\sqrt{n/|\sigma|}}{\|T\|} \leq \frac{1}{\|T\|s} = \delta/s \leq c^{-1}t\log^{3/2}(2/\delta) = t/\alpha.$$

Hence

$$\Big\|\sum_{i\in\sigma} a_i Te_i\Big\| \geq \alpha|\sigma|^{-1/2}\sum_{i\in\sigma}|a_i| \quad \text{for all real numbers } (a_i)$$

as required. ∎

To get the actual invertibility of $T : l_2^n \to X$ one can use the Grothendieck factorization. Remarkably, this step works not only for X being a Hilbert space but for a much larger class of spaces, namely for those of type 2.

Definition 7.6 *A Banach space X has* type 2 *if there exists a constant M such that the inequality*

$$\mathbb{E}\Big\|\sum \varepsilon_i x_i\Big\| \leq M\Big(\sum \|x_i\|^2\Big)^{1/2}$$

holds for all finite sequences of vectors (x_i) in X. The minimal possible constant M is called the type 2 constant of X *and is denoted by $T_2(X)$.*

An important example of spaces that have type 2 are all L_p-spaces ($2 \le p < \infty$) and their subspaces.

Lemma 7.7 (Grothendieck Factorization, see e.g. [LT] 15.4) *Let $S : E \to \mathbb{R}^m$ be a linear mapping, where E is a Banach space of type 2. Then there exists a subset η of $\{1, \ldots, n\}$ of size $|\eta| \ge m/2$ and such that*

$$\|P_\eta S\|_{E \to l_2^m} \le C\, T_2(X)\, m^{-1/2} \|S\|_{E \to l_1^m}$$

where P_η is the coordinate projection in $\mathbb{R}^m$ onto $\mathbb{R}^\eta$.

Applying this lemma to the inverse of T on its range, we obtain

Corollary 7.8 (Restricted Invertibility under type 2) *Let $T : l_2^n \to X$ be a linear operator with $\ell(T) \ge \sqrt{n}$, where X is a Banach space of type 2. Let $\alpha = c \log^{-3/2}(2\|T\|)$. Then there exists a subset σ of $\{1, \ldots, n\}$ of size $|\sigma| \ge \alpha^2 n / \|T\|^2$ and such that*

$$\|Tx\| \ge \alpha\, T_2(X)^{-1} \|x\| \quad \text{for all } x \in \mathbb{R}^\sigma.$$

For $X = l_2^n$ this recovers the Bourgain-Tzafriri Theorem up to the logarithmic factor α.

Proof. By Corollary 7.5, the operator T is invertible on the subspace $E = T(\mathbb{R}^\sigma)$ of X, and its inverse $S = T^{-1} : E \to \mathbb{R}^\sigma$ has norm $\|S\|_{E \to l_1^\sigma} \le \alpha^{-1} |\sigma|^{1/2}$. By the Grothendieck factorization, we find a subset $\eta \subset \sigma$ of size $|\eta| \ge \frac{1}{2}|\sigma|$ and such that

$$\|P_\eta S\|_{E \to l_2^\eta} \le C \alpha^{-1} T_2(X).$$

This means that $\|Tx\|_X \ge c\alpha\, T_2(X)^{-1} \|x\|_{l_2^n}$ for all $x \in \eta$. ■

Finally, the general Principle of the Restricted Invertibility rewritten in geometric terms gives a result related to Dvoretzky's Theorem.

Corollary 7.9 *Let K be a symmetric convex body in $\mathbb{R}^n$ containing B_2^n. Let $M = M_K \log^{-3/2}(2/M_K)$. Then there exists a subset σ of $\{1, \ldots, n\}$ of size $|\sigma| \ge cM^2 n$ and such that*

$$cM(K \cap \mathbb{R}^\sigma) \subset \sqrt{|\sigma|} B_1^\sigma. \tag{7.4}$$

Here B_1^σ denotes the unit ball of l_1^σ.

Proof. We apply the general Principle of Restricted Invertibility in the space $X = (\mathbb{R}^n, \| \cdot \|_K)$. We have $\ell(id : l_2^n \to X) = \sqrt{n} M_K$ (see [TJ] (12.7)) and $\|id : l_2^n \to X\| \le 1$ because K contains B_2^n. Hence for the operator $T = c(M_K)^{-1} id : l_2^n \to X$ we have $\ell(T) \ge \sqrt{n}$ and $\|T\| \le C/M_K$. The application of Corollary 7.5 completes the proof. ■

A link to Dvoretzky Theorem is provided by a result of Kashin [K 77] (see also [S]) that the cross-polytope $\sqrt{k}B_1^k$ has a Euclidean section of proportional dimension. Precisely, there exists a subspace E in $\mathbb{R}^k$ of dimension at least $k/2$ and such that

$$(B_2^k \cap E) \subseteq (\sqrt{k}B_1^k \cap E) \subseteq C(B_2^k \cap E).$$

Actually, a random subspace E taken uniformly in the Grassmanian satisfies this with probability at least $1 - e^{-ck}$.

Taking such random section of both sides of (7.4) we get $M(K \cap E) \subseteq C(B_2^n \cap E)$, which recovers the second inclusion in Dvoretzky's Theorem up to a logarithmic factor. The novelty of (7.4) is that the section is coordinate. This might be important for future applications.

Remark. Corollary 7.9 may fail for any set of size $|\sigma| \asymp M^2 n$, even though it must hold for some larger set. Indeed, for $K = a\sqrt{n}B_1^n$ with some large parameter a we have $M \sim a^{-1}\log^{-3/2} a$. Any set σ for which Corollary 7.9 holds satisfies $\sqrt{|\sigma|}B_1^\sigma \supseteq ca^{-1}(\log^{-3/2} a)(a\sqrt{n}B_1^n \cap \mathbb{R}^\sigma) = c(\log^{-3/2} a)\sqrt{n}B_1^\sigma$, so $|\sigma| \geq (\log^{-3/2} a)n$. This is much larger than $M^2 n \asymp a^{-2}(\log^{-3} a)n$. In particular, Corollary 7.9 fails for any set of size $|\sigma| \sim M^2 n$.

Proof of Theorem 7.4. By a slight perturbation we may assume that the vectors x_i are linearly independent, and by applying appropriate linear transformation we may further assume that $X = (\mathbb{R}^n, \|\cdot\|_K)$ where K is a symmetric convex body in $\mathbb{R}^n$ and that $x_i = e_i$, the canonical vector basis in $\mathbb{R}^n$. We then rewrite the assumptions as $\frac{1}{\sqrt{n}}B_2^n \subseteq K$, $\mathbb{E}\|g\|_K \geq \delta n$. Then for the polar body $A = K^\circ = \{x \in \mathbb{R}^n : \langle x, y\rangle \leq 1 \ \forall y \in K\}$ we have

$$A \subseteq \sqrt{n}B_2^n, \quad E := \mathbb{E}\sup_{x \in A}\sum_{i=1}^n g_i x(i) \geq \delta n. \tag{7.5}$$

Although Theorem 6.5 can be used to estimate E, we will need to have some control on the upper limit in the integral (6.9). This can be done as follows. Noting that $E(A) = n^{-1/2}E$, we bound the expectation in (7.5) by Dudley's entropy inequality (6.10):

$$E \leq C\sqrt{n}\int_{cE/n}^1 \sqrt{\log N(A, tD)}\, dt \tag{7.6}$$

where $D = \mathrm{Ball}(L_2^n) = \sqrt{n}B_2^n$. The upper limit in the integral is 1 because $A \subseteq D$, so the integrand vanishes for $t > 1$. By Theorem 4.1,

$$N(A, tD) = N(t^{-1}A, D) \leq \Sigma(Ct^{-1}A)^2.$$

Since $Ct^{-1}A \subseteq Ct^{-1} \cdot \mathrm{Ball}(L_2^n)$, Lemma 4.7 gives

$$\Sigma(Ct^{-1}A) \le \Big(\frac{Cn}{t\,v(t)}\Big)^{Cv(t)},$$

where $v(t) = v(Ct^{-1}A)$. Hence

$$\log N(A, tD) \le Cv(t) \log\Big(\frac{Cn}{t\,v(t)}\Big).$$

Using this in Dudley's entropy inequality (7.6), we obtain

$$E \le C\sqrt{n} \int_{cE/n}^{1} \sqrt{v(t) \log\Big(\frac{Cn}{t\,v(t)}\Big)}.\ dt$$

Let $s(t)^2 = v(t)/n$. Since $s(t) \le 1$ and $E \ge \delta n$, we have

$$c\delta \le \int_{c\delta}^{1} s(t) \sqrt{\log\Big(\frac{1}{t\,s(t)}\Big)}\ dt.$$

Comparing the integrand to that of

$$\log(1/c\delta) = \int_{c\delta}^{1} \frac{1}{t}\ dt$$

we conclude that there exists a number $c\delta \le t \le 1$ such that

$$s(t) \sqrt{\log\Big(\frac{1}{t\,s(t)}\Big)} \ge \frac{c\delta}{t \log(1/c\delta)}.$$

Multiplying both sides by t, we obtain

$$t\,s(t) \ge \frac{c\delta}{\log(1/c\delta)} \Big/ \sqrt{\log\Big(\frac{\log(1/c\delta)}{c\delta}\Big)} \ge \frac{c\delta}{\log^{3/2}(2/\delta)}.$$

It remains to interpret $v(t)$. By the symmetry of A, $v(t)$ is the maximal rank of a coordinate projection P in $\mathbb{R}^n$ such that $P(Ct^{-1}A) \supseteq P(\frac{1}{2}B_\infty^n)$. Let $\mathbb{R}^\sigma$ be the range of P; then $|\sigma| = v(t) = s(t)^2 n$. By duality, the inclusion above is equivalent to $C^{-1}tK \cap \mathbb{R}^\sigma \subseteq 2B_1^n$. Equivalently, $\|x\|_K \ge C^{-1}t\|x\|_{l_1^n}$ for all $x \in E$. This is precisely the conclusion (7.3). The proof is complete. ■

Remark. Although the first assumption in (7.2) is rather nonrestrictive, it can further be weakened. Tracing where it was used in the proof (in Lemma 4.7) we see that only "average" volumetric properties of K matter. We leave details to the interested reader.

References

[A] S. Alesker, *Remark on Szarek-Talagrand theorem*, Combin. Probab. Comput. 6 (1997), 139–144

[AB] M. Anthony, P.L. Bartlett, *Neural Network Learning, Theoretical Foundations*, Cambridge University Press, 1999.

[ABCH] N. Alon, S. Ben-David, N. Cesa-Bianchi, D. Hausser, *Scale sensitive dimensions, uniform convergence and learnability*, Journal of the ACM 44 (1997), 615–631

[BKP] P.L. Bartlett, S.R. Kulkarni, S.E.Posner, *Covering numbers for real-valued function classes*, IEEE Trans. Inform. Theory 43 (1997) 1721–1724

[BL] P.L. Bartlett, P. Long, *Prediction, learning, uniform convergence, and scale-sensitive dimensions*, Eighth Annual Workshop on Computational Learning Theory (COLT) (Santa Cruz, CA, 1995), J. Comput. System Sci. 56 (1998), 174–190

[BT 87] J. Bourgain, L. Tzafriri, *Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis*, Israel J. Math. 57 (1987), 137–224

[BT 91] J. Bourgain, L. Tzafriri, *On a problem of Kadison and Singer*, J. Reine Angew. Math. 420 (1991), 1–43

[CFG] H.T.Croft, K.J.Falconer, R.K.Guy, *Unsolved problems in geometry*, Problem Books in Mathematics. Unsolved Problems in Intuitive Mathematics, II. Springer-Verlag, New York, 1991

[Du 78] R.M. Dudley, *Central limit theorems for empirical measures*, Ann. Probab. 6 (1978), 899–929.

[Du 99] R.M. Dudley, *Uniform central limit theorems*, Cambridge University Press, 1999.

[E] J. Elton, *Sign-embeddings of* l_1^n, Trans. AMS 279 (1983), 113–124

[GZ] E. Gine, J. Zinn, *Gaussian characterization of uniform Donsker classes of functions*, Annals of Probability 19 (1991), 758–782

[Ha] D. Haussler, *Generalizing the PAC model: sample size bounds from metric dimension-based uniform convergence results*, Proceedings of the 30th Annual Symposium on the Foundations of Computer Science, 1989

[HL] D. Haussler, P. Long, *A generalization of Sauer's lemma*, J. Combin. Theory Ser. A 71 (1995), 219–240

[H] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963), 13–30

[K 77] B. Kashin, *Sections of some finite dimensional sets and classes of smooth functions*, Izv. Acad. Nauk SSSR 41 (1977), 334-351 (Russian)

[K 85] B. Kashin, *On a certain isomorphism on* $L^2(0,1)$, Comptes Rendus Acad. Bulgare des Sciences 38(1985), 1613–1615 (Russian)

[KS] M. Kearns, R. Schapire, *Efficient distribution-free learning of probabilistic concepts*, J. Comput. System Sci. 48 (1994), 464–497

[LT] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer, 1991

[M 01] J. Matoušek, *On directional convexity*, Discrete Comput. Geom. 25 (2001), 389–403

[M 02] J. Matoušek, *Lectures on discrete geometry*, Graduate Texts in Mathematics, 212, Springer-Verlag, New York, 2002

[MS] V. Milman and G.Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Math. 1200, Springer, 1986

[MV 02] S. Mendelson, R. Vershynin, *Entropy, Combinatorial dimensions and Random Averages*, in: Computational Learning Theory (15th Annual Conference on Computational Learning Theory), edited by J.Kivinen and R.Sloan, Springer 2002, 14–28

[MV 03] S. Mendelson, R. Vershynin, *Entropy and the combinatorial dimension*, Inventiones Mathematicae 152 (2003), 37–55

[Pa 82] A. Pajor, *Plongement de ℓ_1^k dans les espaces de Banach*, Sèminaire d'Analyse Fonctionelle, PAris VII, 1982-83, à paraître

[Pa 85] A. Pajor, *Sous espaces ℓ_1^n des espaces de Banach*, Hermann, Paris, 1985

[Pi] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics 94, Cambridge University Press, 1989.

[Po] D. Pollard, *Convergence of Stochastic Processes*, Springer Verlag, 1984

[S] S. Szarek, *On Kashin's almost Euclidean orthogonal decompositions of l_1^n*, Bull. Acad. Polon. Sci. 26 (1978), 691–694

[ST] S. Szarek, M. Talagrand, *On the convexified Sauer-Shelah theorem*, J. Combinat. Theory, Ser. B 69 (1997), 183–192

[T 87] M. Talagrand, *The Glivenko-Cantelli problem*, Annals of Probability 15 (1987), 837–870

[T 92] M. Talagrand, *Type, infratype, and Elton-Pajor Theorem*, Inventiones Math. 107 (1992), 41–59

[T 96] M. Talagrand, *The Glivenko-Cantelli problem, ten years later*, J. Theoret. Probab. 9 (1996), 371–384

[T 02] M. Talagrand, *Vapnik-Chervonenkis type conditions and uniform Donsker classes of functions*, Ann. Probab. 31 (2003) 1565–1582

[T 03] M. Talagrand, *Vapnik-Chervonenkis type conditions and uniform Donsker classes of functions*, Ann. Prob. 31 (2003), 1565–1582

[TJ] N. Tomczak-Jaegermann, *Banach-Mazur distances and finite dimensional operator ideals*, Pitman, 1989

[VC 68] V. Vapnik, A. Chervonenkis, *Uniform convergence of frequencies of occurence of events to their probabilities*, Dokl. Akad. Nauk SSSR 181 (1968), 781–783 (Russian), Sov. Math. Doklady 9 (1968), 915–918 (English translation)

[VC 71] V. Vapnik, A. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory Probab. Appl. 16 (1971), 264–280

[VC 81] V. Vapnik, A. Chervonenkis, *Necessary and sufficient conditions for the uniform convergence of empirical means to their expectations*, Theory Probab. Appl. 3 (1981), 532–553

M. Rudelson, Department of Mathematics, University of Missouri, Columbia, MO 65211, USA; e-mail: rudelson@math.missouri.edu

R. Vershynin, Department of Mathematics, University of California, Davis, CA 95616, USA; e-mail: vershynin@math.ucdavis.edu