ENTROPY METHODS

Shiri Artstein

August 4, 2004

Acknowledgements

First and foremost I would like to thank my supervisor and teacher Professor Vitali Milman. Professor Milman's help, guidance, support and dedication, were invaluable. His contribution of ideas, continuous advice and encouragement, and his kindness, have made being a student of his a real experience. His contribution to my understanding of mathematics and to my research cannot be overestimated.

I would like to thank also my collaborators, Prof. Franck Barthe, Dr. Assaf Naor, Prof. Staszek Szarek and Prof. Nicole Tomczak-Jaegermann for their contributions to my work and for fruitful discussions. I would like to thank especially Prof. Keith Ball whose influence on my research and to my progress in mathematics goes beyond what is reflected in our joint work.

Part of the research was carried out while I was visiting, and enjoying the warm hospitality of, University College London, University Paris 6 (Université Pierre et Marie Curie) and University Paris 7 (Marne-la-Vallée).

Abstract

The topics of this thesis are at the meeting point of Analysis and Geometry. In this field of Asymptotic Functional Analysis one studies asymptotic phenomena in finite dimensional linear spaces as the dimension of the space grows to infinity. In particular, the limit behavior as n tends to infinity of certain properties and parameters related to subsets of n-dimensional spaces are examined. The properties of convex bodies play a major role. The field has evolved from the field of Geometric Functional Analysis, providing a new approach to some central problems. The theory, for example, enables us to investigate infinite dimensional spaces via their finite dimensional sections or projections.

This thesis is concentrated around two main directions, which are connected: metric entropy and informational theoretic entropy.

The main part of the results consists of the solution of two open problems, each more than 30 years old. The first (joint work with V. Milman and S. Szarek) is the solution of the Duality of Entropy Numbers Conjecture, in the case where one of the two spaces involved is euclidean. This conjecture was posed by Pietsch in 1972, as a consequence of "experimental" data regarding the behavior of sequences of entropy numbers for operators and their duals, which he computed. The case where one of the spaces is euclidean has always been considered as the main case and was investigated by Bourgain, König, Milman, Pajor, Pisier, Schütt, Szarek, Tomczak-Jaegermann and others. The proof of the conjecture reveals a strong and surprising connection between the geometric structure of a space and its dual. The dissertation includes also generalizations of this theorem (joint work with V. Milman, S. Szarek and N. Tomczak-Jaegermann) to the case where, for example, one of the two spaces has some non trivial type $p > 1$. To this end we introduce a new concept of "convex separation", which we investigate, and connect with usual covering numbers (which are the inverse of entropy numbers).

The second long standing problem, the solution of which I present in this

thesis, dates back to Shannon in the 50s and was stated explicitly as a conjecture by Lieb in the 70s. This is joint work with K. Ball, F. Barthe and A. Naor. Our result deals with Shannon's informational theoretic entropy and states that the entropy of a random system of normalized sums of n independent identically distributed random variables is an increasing function of n. The fact that as n doubles the relevant entropy grows was proved by Shannon, but even the fact that the entropy of the normalized sum of 3 copies is no less than that of 2 was an open problem. Another result we prove is an estimate for the rate of increase of entropy in random systems under the assumption of a spectral gap; the result improves upon previously known results.

The thesis includes several more results which also deal with phenomena in high dimensions. I present an asymptotically exact computation of the area of some special sets on the sphere, namely, the neighborhoods of subspaces. This result serves as a technical tool for covering number results but it is also of independent interest and has other applications. We present results concerning the decay of diameter of convex bodies under a certain type of small rank operator. We conclude with further remarks on covering numbers. First we will see an interesting phase transition behavior of entropy numbers with respect to projections, secondly we define the concept of 'weighted covering' and prove some basic properties for this concept; finally, we remark on coverings in which rotates of the covering body are allowed rather than just translates.

Contents

Preface

In this thesis we display results regarding two different kinds of entropy. One is the informational-theoretic **Shannon's entropy**, which for a continuous random variable with density f is defined by $-\int f \ln f$. Another is **Metric entropy**, which is defined for a compact linear operator $u : X \to Y$ between two Banach spaces, and associates to u the sequence $e_k(u)$ (called "entropy numbers of u") which quantifies in some sense the compactness of u. Both notions are related since both measure the disorder of an object, hence the name "entropy".

In the first chapter we describe the results connected with Metric entropy. Our main result in this direction, which is joint work with V. Milman and S. Szarek, is the solution of the Duality of Entropy Numbers Conjecture, in the case where one of the two Banach spaces involved is a Hilbert space. This conjecture was posed by Pietsch in 1972, as a consequence of his "experimental" work regarding the behavior of entropy number sequences for specific operators and their duals, which he was able to compute. The case when one of the spaces involved is euclidean has always been considered the main case (it is almost the *only* case used in applications, and has a clear geometric meaning), and was investigated by Bourgain, König, Milman, Pajor, Pisier, Schütt, Szarek, Tomczak-Jaegermann and others, see e.g. [BPST], [KöM], [MSz2], [PaTo2], [Sch] and [Tom]. The proof of the conjecture reveals a strong and surprising connection between the geometric structure of a space and its dual. The proof is based on the papers [AMS1], [AMS2] and [AMS3]. We also prove generalizations of this theorem to the case where, for example, one of the two spaces has some non trivial type $p > 1$. This is joint work with V. Milman, S. Szarek and N. Tomczak-Jaegermann, and appears in [AMST]. To this end we introduce a new concept of "convex separation", which we investigate, and connect with usual covering numbers (which are the inverse of entropy numbers). Chapter 1 of this thesis is dedicated to covering and entropy numbers results.

In the second chapter of this thesis we explain results concerning Shannon's entropy. Our first main result in this direction is the solution of a conjecture which dates back to Shannon in the 50s and was stated explicitly as a conjecture by Lieb in the 70s. This is joint work with K. Ball, F. Barthe and A. Naor. Our result states, in particular, that the entropy of a random system of normalized sums of n independent identically distributed random variables is an increasing function of n. The fact that as n doubles the relevant entropy increases was proved by Shannon, but even the fact that the entropy of the normalized sum of 3 copies is no less than that of 2 was unknown. This result has appeared in [ABBN1]. In our second main result in this direction, which we prove with related methods, we estimate the rate of increase of entropy in random systems under the assumption of a spectral gap. Our result improves upon previously known estimates, and has appeared in [ABBN2]. Chapter 2 is devoted to presenting these questions and results.

The additional results displayed in this thesis also deal with phenomena in high dimensions, and are described in Chapter 3. Some of them have appeared in the papers [Art1] and [Art2]. In Section 3.2 we will see an asymptotically exact computation of the area of some special sets on the sphere, namely, the neighborhoods of subspaces. This estimate serves as a technical tool for entropy and covering numbers results in Chapter 1 but it is also of independent interest and has other applications. In Section 3.3 we show results concerning the decay of diameter of convex bodies under a certain type of small rank operator, the proof of which also includes interesting covering methods. In Section 3.4 we have gathered some further remarks regarding covering numbers. We will see an interesting phase transition behavior of entropy numbers with respect to projections, we will define the concept of 'weighted covering' and prove some basic properties for this concept; finally we will remark on covering where rotated bodies are allowed rather than just translates.

Chapter 1

Metric Entropy

1.1 Introduction

In this chapter we describe our results concerning Metric Entropy and Covering Numbers. The covering number of a convex body K by a convex body T, denoted $N(K,T)$, is defined as the minimal number of translates of T needed to cover K. If T is the unit ball of one space, and K is the image of the unit ball of another space under some given compact linear operator u, the number $N(K,tT)$ as a function of the parameter t quantifies in some sense the compactness of u. (Metric entropy in operator theory is the inverse of this function, precise definitions below). The computation of entropy numbers is usually extremely difficult. However, entropy numbers arise very naturally in the solution of many problems in analysis and in probability, for example in the study of gaussian processes; understanding their behavior is an important goal.

A well known 30 year old conjecture in this field is the duality of entropy numbers conjecture, which in the language of covering numbers states that in an appropriate sense, as functions of t, the expressions $N(K,tT)$ and $N(T^\circ,tK^\circ)$ are equivalent (see Theorem 1.2.1 below), where K° and T° denote the polar bodies of K and T respectively. The most important case (which is the one appearing in applications to probability theory, ergodic theory, learning theory and other fields) is when one of the two bodies is assumed to be the euclidean ball. In the language of operators, this means that either the domain or the range of the operator u is a Hilbert space. This case has been studied by many mathematicians in the field and recently, jointly with V. Milman and S. Szarek, we proved the duality conjecture in this case. After giving precise definitions and the statement of the theorem in Section 1.2, and useful easy facts in Section

1.3, we demonstrate this theorem's proof in Section 1.4.

The case of two general convex bodies, or equivalently of an operator between two general spaces, is still under investigation, however we have recently managed to prove a generalization of the theorem to the case in which one of the spaces has some non trivial type $p > 1$ (and the constants involved in the theorem will depend both on the type p and on the type constant of the space). These include, in particular, all ℓ_p- and L_p-spaces (classical or non-commutative) for $1 < p < \infty$, and similarly in all uniformly convex and uniformly smooth spaces. This is joint work with V. Milman, S. Szarek and N. Tomczak-Jaegermann. To prove this generalization we needed to define a new concept - similar to that of covering numbers - which we call "convex separation" or "convexified packing" number. Section 1.5 is devoted to this new concept, its properties, and the way it can be used to give duality theorems.

This chapter includes results from the papers [AMS1] [AMS2] [AMS3] and [AMST].

1.2 Definitions, More Introduction, Statement of the Main Result

For two convex bodies K and T in $\mathbb{R}^n$, the covering number of K by T, denoted $N(K,T)$, is defined as the minimal number of translates of T needed to cover K, that is,

$$N(K,T) = \min\{N : \exists\, x_1 \dots x_N \in \mathbb{R}^n,\ K \subset \bigcup_{i\le N} x_i + T\}.$$

We denote by $K^\circ := \{u \in \mathbb{R}^n : \sup_{x\in K}\langle x,u\rangle \le 1\}$ the polar body of K, and by D the euclidean unit ball in $\mathbb{R}^n$. In this chapter we prove the following duality result for covering numbers.

Theorem 1.2.1 (Main theorem) *There exist two universal constants α and β such that for any dimension n and any convex body $K \subset \mathbb{R}^n$, symmetric with respect to the origin, one has*

$$N(D, \alpha^{-1}K^\circ)^{\frac{1}{\beta}} \le N(K,D) \le N(D,\alpha K^\circ)^{\beta} \tag{1.2.1}$$

The best constant β that our approach yields is $\beta = 2+\varepsilon$ for any $\varepsilon > 0$, with $\alpha = \alpha(\varepsilon)$.

This theorem establishes a strong connection between the geometry of a set and its polar or, equivalently, between a normed space and its dual. Notice that since the theorem is true for any K, we can actually infer that for any $t > 0$

$$\beta^{-1} \log N(D, \alpha^{-1} tK^\circ) \leq \log N(K, tD) \leq \beta \log N(D, \alpha tK^\circ). \tag{1.2.2}$$

(For definiteness, above and in what follows all logarithms are to the base 2.) The quantity $\log N(K, tT)$ has a clear information-theoretic interpretation: it is the complexity of K, measured in bits, at the level of resolution t with respect to the metric associated with T (e.g., euclidean if $T = D$). Accordingly, (1.2.2) means that the complexity of K in the euclidean sense is controlled by that of the euclidean ball with respect to $\|\cdot\|_{K^\circ}$ (the gauge of K°, i.e., the norm whose unit ball is K°), and vice versa, at *every* level of resolution. While it is clear that these complexities should be related, the universality of the link that we establish is somewhat surprising.

In addition to the immediate information-theoretic ramifications, covering numbers appear in many other areas of mathematics. For example, both quantities $N(K, tD)$ and $N(D, tK^\circ)$ enter the theory of Gaussian processes (see, e.g., [Du] and [KuLi], or the survey [Led]) and our results transform some conditional statements into theorems (see, e.g., [LiLi]). More precisely, a Banach space X with unit ball K was said to satisfy the $\mathcal{D}$-property if the conclusion in the statement of Theorem 1.2.1 held for K, and theorems were proved under the condition that the space involved has the $\mathcal{D}$-property. Our Theorem 1.2.1 says that *every* Banach space has the $\mathcal{D}$-property.

Theorem 1.2.1 resolves an old problem, going back to Pietsch ([Pie], p. 38) and referred to as the "duality conjecture for entropy numbers," in a special yet most important case. The problem can be stated in terms of covering numbers in the following way (below and in what follows we shall abbreviate "symmetric with respect to the origin" to just "symmetric").

Conjecture 1.2.2 (Duality Conjecture) *Do there exist two numerical constants $a, b \geq 1$ such that for any dimension n, and for any two symmetric convex bodies K and T in $\mathbb{R}^n$ one has*

$$\log N(T^\circ, aK^\circ) \leq b \log N(K, T), \tag{1.2.3}$$

where A° denotes the polar body of A ?

By affine invariance of the problem Theorem 1 actually verifies this conjecture in the case where one of the two bodies is an ellipsoid. In the special case where *both* bodies are ellipsoids it is easy to show that there is equality in (1.2.3), with $a = b = 1$ (we state it as Fact 1.3.2 below, and it follows directly from Fact 1.3.1).

This conjecture originated in operator theory, and so we restate it below in the language of entropy numbers of operators. For two Banach spaces X and Y, with unit balls $B(X)$ and $B(Y)$ respectively, and for a linear operator $u : X \to Y$, the k^{th} entropy number of u is defined by

$$e_k(u) := \inf\{\varepsilon : N(uB(X), \varepsilon B(Y)) \leq 2^{k-1}\}.$$

(In fact, above and in what follows k does not need to be an integer.) So, for example, $e_1(u) = \|u\|_{op}$ (the operator norm), and one can easily see that $e_k(u) \to 0$ as $k \to \infty$ if and only if u is a compact operator. Therefore the two sequences $(e_k(u))$ and $(e_k(u^*))$ always begin with the same number $\|u\|_{op} = \|u^*\|_{op}$, and $e_k(u) \to 0$ if and only if $e_k(u^*) \to 0$. Since the sequence $(e_k(u))$ can be thought of as quantifying the compactness of the operator u, it seems natural to ask to what extent do $(e_k(u))$ and $(e_k(u^*))$ behave similarly. This is the context in which the duality conjecture was originally formulated, and it read as follows.

(Duality Conjecture in the language of entropy numbers) *Do there exist numerical constants $a, b \geq 1$, such that for any two Banach spaces X and Y, any linear operator $u : X \to Y$ and any natural number k, one has*

$$e_{bk}(u^*) \leq a e_k(u) \ ?$$

The two formulations are equivalent since considering the entropy numbers of the dual operator $u^* : Y^* \to X^*$ means covering $(B(Y))^\circ$ with (translates of) $\varepsilon(B(X))^\circ$, and since one can restrict oneself to bodies which are convex hulls of a finite number of points and thus lie in a finite dimensional space. This is formulated explicitly in Observation 1.4.7 below. Thus, Theorem 1.2.1 verifies the duality conjecture (when expressed in terms of entropy numbers) for the case in which one of the two spaces, either the domain or the range of the operator u, is a Hilbert space.

Some special cases of the problem have been studied before, and some particular results were established, see e.g., [Art1], [AMS1], [BPST], [GKS], [KöM], [MSz1], [MSz2], [PaTo2], [Pis2], [Pis3], [Sch], [Tom]. Two of the above have special relevance to our approach: Firstly [KöM], which shows the duality of entropy numbers when the rank of the operator is (at most) comparable to the logarithm of the covering number. This result is stated in Lemma 1.4.6 and a variant of it is proved in Corollary 1.5.13 in Section 1.5.5.3. Secondly [Tom], which demonstrates a form of duality involving some measures of the size of the *entire* sequences $(e_k(u)), (e_k(u^*))$.

Remark on notation: Unless otherwise stated, above and in what follows all constant appearing are universal (notably independent of the dimension and of the particular convex body or the operator that is being considered). If a constant c depends on some parameter θ, we will indicate that by writing $c(\theta)$. Secondly, the same letters c, C can denote different constants in this chapter (and thesis).

1.3 Helpful facts regarding covering numbers

In this section we gather some of the basic well known properties of covering numbers which will be constantly used in this chapter. Their proofs are straightforward.

Fact 1.3.1 *For every $A, B \subset \mathbb{R}^n$ and every invertible linear operator $v : \mathbb{R}^n \to \mathbb{R}^n$ one has $N(A,B) = N(vA, vB)$.*

Fact 1.3.2 *For any two ellipsoids $\mathcal{E}_1, \mathcal{E}_2$ one has $N(\mathcal{E}_1, \mathcal{E}_2) = N((\mathcal{E}_2)^\circ, (\mathcal{E}_1)^\circ)$.*

Fact 1.3.3 *For every A, B and C it is true that $N(A,B) \le N(A,C)N(C,B)$.*

Fact 1.3.4 *For every A, B and C it is true that $N(A + C, B + C) \le N(A,B)$*

Fact 1.3.5 *For every A and B it is true that $N(A, 2(B \cap A)) \le N(A,B)$*

Fact 1.3.6 *If D is a euclidean ball then $N(A, D \cap 2A) = N(A,D)$*

Define the packing number $M(K,T)$ as the maximal cardinality of a T-separated set in K, that is,

$$M(K,T) = \max\{N : \exists x_1, \dots, x_N \in K \quad \text{s.t. } \forall j \neq i, x_j \notin (x_i + T).\}$$

This number is closely related to the covering number, and it is easily verified that

Fact 1.3.7 $M(A, 2B) \le N(A,B) \le M(A,B)$.

1.4 Proof of the duality Theorem 1.2.1

In this section we describe *two* proofs of Theorem 1.2.1. Both rely on the same basic ingredients, although in each the 'recipe' is slightly different. The first is conceptually simpler, and also contains steps which are easier to generalize.

However, it does not give as good constants as does the second method, which is inherently euclidean and thus harder to generalize. In Section 1.4.1 we prove a proposition (Proposition 1.4.1, which we call the 'reduction proposition') which can be considered the main step in the first-method proof of Theorem 1.2.1; this proposition will also be used in the generalizations of Theorem 1.2.1 in Section 1.5. In Section 1.4.2 we complete the first proof. In Section 1.4.3 we give the alternative proof, which provides the best constants to date.

1.4.1 Reduction Proposition

The following proposition enables us, while trying to prove a duality theorem for two general convex bodies K and T (not requiring, yet, that one of them be a euclidean ball), to restrict our attention to the special case in which $K \subset 4T$.

Proposition 1.4.1 *Let T be a convex symmetric body in a euclidean space such that, for some constants $c, C > 0$, the following holds: if K is a convex symmetric body with $K \subset 4T$, then*

$$N(K,T) \le N(T^\circ, cK^\circ)^C.$$

Then, for some other constants $c', C' > 0$ (depending only on c, C) and **any** *convex symmetric body K*

$$N(K,T) \le N(T^\circ, c'K^\circ)^{C'}.$$

Dually, if K is fixed and the hypothesis holds for all T's verifying $K \subset 4T$, then the assertion holds for **any** *T.*

There is no special magic to the number 4 in this proposition, and the same result holds if, instead, we assume in the hypothesis that it holds true for all $K \subset RT$, where $R > 2$ is some fixed constant. We use 4 only to simplify notation. Moreover, in the 'Dually' part, it is enough to assume that the condition holds for all $K \subset RT$ for some fixed $R > 1$, and if the body T is the euclidean ball then $R > 1$ suffices in the first part as well.

This proposition is proved by an iterative scheme. We need two lemmas. The first lemma is based on a simple geometric iteration procedure.

Lemma 1.4.2 (Iterating) *For any two symmetric convex bodies $K, T \subset \mathbb{R}^n$ and any integer $s \ge 1$,*

$$N(K,T) \le N(K, 2^sT) \prod_{j=0}^{s-1} N(2^{1-j}K \cap 4T, T) \tag{1.4.1}$$

and

$$N(K,T) \le N(K,2^sT)\prod_{j=0}^{s-1} N(K,\mathrm{conv}(2^{j-1}T \cup \frac{1}{4}K)) \qquad (1.4.2)$$

(where '$\mathrm{conv}(A)$*' denotes the convex hull of the set* A*).*

Proof For (1.4.1) consider the following inequality

$$\begin{aligned} N(K,2^jT) &\le N(K,2^{j+2}T\cap 2K)N(2K\cap 2^{j+2}T,2^jT) \\ &\le N(K,2^{j+1}T)N(2K\cap 2^{j+2}T,2^jT) \end{aligned}$$

which follows from Facts 1.3.3 and 1.3.5. Iterating this inequality from $j=0$ to s, and dividing both entries in the second term by 2^j, gives (1.4.1). For (1.4.2), consider the following inequality which follows from Fact 1.3.3

$$N(K,2^jT) \le N(K,\mathrm{conv}(2^{j-1}T\cup\frac{1}{4}K))N(\mathrm{conv}(2^{j-1}T\cup\frac{1}{4}K),2^jT).$$

Changing the convex hull in the second term to the Minkowski sum of the sets (which is bigger and thus harder to cover), using Fact 1.3.4, and multiplying both entries by 2^{j+1}, leads to

$$N(K,2^jT) \le N(K,\mathrm{conv}(2^{j-1}T\cup\frac{1}{4}K))N(K,2^{j+1}T).$$

Repeating the above argument $(s-1)$ times more yields (1.4.2). The proof of Lemma 1.4.2 is thus complete. □

For each factor in these products, the body that is being covered is included in 4 times the covering body, and so we will be able to use the assumption of the proposition. (We will take s to be the smallest integer such that $N(K,2^sT)=1$). We will thus have a bound of the original covering number by a product of dual covering numbers. To collapse this product of many numbers to the product of just a few, we need the following super-multiplicativity inequalities for covering numbers. We omit the notation $\mathrm{conv}(\cdot)$ in (1.4.4), and the unions in inequality (1.4.4) should be understood as the convex hulls of the unions.

Lemma 1.4.3 *Let* $A > a > 3B > 3b$*. Then for any two symmetric convex bodies* K *and* T*,*

$$\begin{aligned} N(K\cap AT,aT)N(K\cap BT,bT) &\le N(K\cap AT,\frac{b}{4}T) & (1.4.3)\\ N(T,a\,(K\cup\frac{1}{A}T))N(T,b\,(K\cup\frac{1}{B}T)) &\le N(T,\frac{b}{4}\,(K\cup\frac{1}{A}T)). & (1.4.4) \end{aligned}$$

Proof For the first inequality, first notice that K enters the inequality only via its intersections with T-balls of radii $\leq A$, so we may as well assume that $K = K \cap AT$ to begin with. Denote $N_1 = N(K, aT)$ and $N_2 = N(K \cap BT, bT)$. Pick (by Fact 1.3.7) an aT-separated set $x_1, \dots x_{N_1}$ in K and a bT-separated set $y_1, \dots y_{N_2}$ in $K \cap BT$. Define a new set by $z_{i,j} = x_i/2 + y_j/2$. All these points are in K, and there are N_1N_2 of them. We shall show that, in addition, the $z_{i,j}$'s are $\frac{b}{2}T$-separated; this will, by Fact 1.3.7, imply that $N(K, \frac{b}{4}T) \geq N_1N_2$, as required. To show the asserted separation, we denote by $||\cdot||$ the norm associated to the body T, and consider two cases. First, if we look at $||z_{i,j} - z_{i,k}||$, this is simply $||y_j - y_k||/2$ and it exceeds $b/2$. On the other hand if $k \neq i$, then by the triangle inequality $||z_{i,j} - z_{k,l}|| \geq ||x_i - x_k||/2 - ||y_j - y_l||/2$, and using the fact that the y_i's are in BT we see that these quantities are greater than $a/2 - B$, which in turn exceeds $\frac{b}{2}$. This completes the proof of inequality (1.4.3).

For the second inequality in the lemma, we this time notice firstly that K enters the inequality only via the convex hull of its union with T-balls of radii $\geq \frac{1}{A}$, so we may as well assume that $K = K \cup \frac{1}{A}T$ to begin with. Denote $N_1 = N(T, aK)$ and $N_2 = N(T, b\,\mathrm{conv}(K \cup \frac{1}{B}D))$. Pick, by Fact 1.3.7, sets $\{x_1, \dots, x_{N_1}\}$ and $\{y_1, \dots y_{N_2}\}$ in T which are respectively aK-separated and $b(\alpha K + \frac{1-\alpha}{B}T)$-separated, where $\alpha = \frac{a}{2a-b} \in (0,1)$ (note that $\alpha K + \frac{1-\alpha}{B}T \subset \mathrm{conv}(K \cup \frac{1}{B}T)$). Define $z_{i,j} = \frac{b}{2a}x_i + (1 - \frac{b}{2a})y_j$. All these points are in T, and there are N_1N_2 of them. As above, it will be enough to show that the $z_{i,j}$'s are $\frac{b}{2}K$-separated, i.e., that whenever $(i,j) \neq (k,l)$, we have

$$z_{k,l} \notin z_{i,j} + \frac{b}{2}K.$$

When looking at $j = l$, this is the same as asking that $\frac{b}{2a}x_k \notin \frac{b}{2a}x_i + \frac{b}{2}K$, which follows from the separation of x_i and x_k. When looking at $j \neq l$ and noticing that $x_i, x_k \in T$, we see that it suffices to show that

$$(1 - \frac{b}{2a})y_j \notin (1 - \frac{b}{2a})y_l + 2\frac{b}{2a}T + \frac{b}{2}K.$$

Under our hypotheses, the above follows from the separation of y_l and y_j. Indeed, $\frac{1}{2}(1 - \frac{b}{2a})^{-1} = \alpha$ just by the definition of α. On the other hand, it is readily verified that the assumption $a > 3B > 2B + b$ implies $\frac{1}{a}(1 - \frac{b}{2a})^{-1} < \frac{1-\alpha}{B}$. The proof is thus complete. □

Remark With more careful arguing, any factor less than $1/2$ instead of $1/4$ can be obtained in the lemma (with then stronger conditions on a, B). Moreover, if we work all along with separated sets instead of covering numbers, then we may arrive at any factor less than 1: for a factor $1 - \varepsilon$ (which corresponds to

$\frac{1}{2} - \frac{\varepsilon}{2}$ for covering numbers), we need the condition $a > 3(\frac{1-\varepsilon}{\varepsilon})B$. This can be used to improve slightly the constants in our main theorem.

Proof of Proposition 1.4.1 Let us first assume that the convex symmetric body T is fixed. The assumption in the proposition is that there exist two constants $c, C > 0$, such that for every L which is a convex symmetric body with $L \subset 4T$, we have

$$N(L,T) \le N(T^\circ, cL^\circ)^C. \tag{1.4.5}$$

Let K by any convex symmetric body. Denote by s the smallest integer such that $N(K, 2^sT) = 1$. Inequality (1.4.1) of Lemma 1.4.2 tells us that

$$N(K,T) \le \prod_{j=0}^{s-1} N(2^{1-j}K \cap 4T, T).$$

We use, for every j, inequality (1.4.5) for the body $L = 2^{1-j}K \cap 4T$ (which is clearly inside $4T$), to get that

$$N(K,T) \le \prod_{j=0}^{s-1} N(T^\circ, c(2^{1-j}K \cap 4T)^\circ)^C,$$

and rewrite it as

$$N(K,T)^{1/C} \le \prod_{j=0}^{s-1} N(T^\circ, c2^{j-1}\mathrm{conv}(K^\circ \cup \frac{1}{2^{j+1}}T^\circ)). \tag{1.4.6}$$

We would now like to collapse this product using (1.4.4) of Lemma 1.4.3, (with the body T° acting as T and K° as K).

However, an additional trick is required since for two neighboring factors in the products the condition of Lemma 1.4.3 does not hold, and so they cannot be "collapsed." For example, for two such factors in (1.4.6) one has $a = c2^{j-1}$ and $B = 2^j$, and so one cannot hope for $a > 3B$. The trick is to split the product into several parts, say l sub-products, for some constant l. We do this by grouping separately the factors corresponding to the j's with some fixed remainder modulo l. If two terms in the product are l terms apart, that is, correspond to some j and $j-l$, then we have $a = c2^{j-1}$ and $B = 2^{j-l+1}$, so as long as $c2^{j-1} > 32^{j-l+1}$, that is, $l > \log\frac{3}{c} + 2$, the two terms are collapsable. We will require a bit more, because we also lose a factor 4 in every use of (1.4.4), but these terms do not accumulate and are each time consumed in the condition of the next step.

More precisely, we denote $l = [\log\frac{12}{c}] + 3$, and rewrite the right hand side of (1.4.6) as (without loss of generality we may assume that $\log l$ divides $s+1$, otherwise we increase s so that it is the case)

$$\prod_{k=0}^{l-1}\left(\prod_{j=0}^{\frac{(s-(l-1))}{l}} N(T^\circ, c2^{(lj+k)-1}\mathrm{conv}(K^\circ \cup \frac{1}{2^{(lj+k)+1}}T^\circ))\right).$$

We will bound from above each of the l products, using (1.4.4). To collapse the last term in the k^{th} product with the one before last in the k^{th} product we need to check that $a = c2^{(s-l+1+k)}$ is greater that $3B = 3 \times 2^{(s-2l+1+k)}$, which clearly holds from the choice of l. We thus bound the k^{th} product by

$$N(T^\circ, \frac{c}{4}2^{(s-2l+k)}\mathrm{conv}(K^\circ \cup \frac{1}{2^{(s-2l+k)+2}}T^\circ)) \times$$
$$\prod_{j=0}^{\frac{(s-(l-1))}{l}-2} N(T^\circ, c2^{(lj+k)-1}\mathrm{conv}(K^\circ \cup \frac{1}{2^{(lj+k)+1}}T^\circ))$$

From here onward all the steps are the same; we just need to make sure at each stage j that

$$\frac{c}{4}2^{(lj+k)-1} \geq 3 \times 2^{(lj-l+k)+1} \tag{1.4.7}$$

As before, this can be rewritten as $2^{l-2} \geq \frac{12}{c}$, which clearly holds for our choice of l. We point out that the factors $1/4$ in $b/4$ do not accumulate, but enter into the quantity a of the next step.

Continuing this way with all the factors of these products, we arrive at

$$N(K,T)^{1/C} \leq \prod_{k=0}^{l-1}\left(N(T^\circ, \frac{c}{4}2^{k-1}\mathrm{conv}(K^\circ \cup \frac{1}{2^{k+1}}T^\circ))\right),$$

which clearly implies

$$N(K,T)^{1/C} \leq N(T^\circ, \frac{c}{8}K^\circ)^l.$$

This completes the proof of the reduction proposition in the case where T is a fixed body. We get the estimate $c' = \frac{c}{8}$ and $C' = Cl = C([\log \frac{12}{c}] + 3)$.

The proof of the second case, when one assumes the condition in the reduction proposition for a fixed body K and all T with $T \supset \frac{1}{4}K$, and deduce it (with different constants) for **all** Ts, is fully analogous. One first uses the bound (1.4.2) instead of (1.4.1) of Lemma 1.4.2, then the assumption in the proposition, and then collapses the product using (1.4.3) instead of (1.4.4). □

1.4.2 Completing the proof of Theorem 1.2.1

To finish the proof in the case where one of the two bodies is the euclidean ball D, we need to combine the reduction Proposition 1.4.1 with the following two simple inequalities, which were first noticed by N. Tomczak-Jaegermann in [Tom], and are in fact themselves duality theorems for bodies K with small diameter.

Lemma 1.4.4 *Let K be a convex symmetric body, and D the euclidean unit ball.*

(1) If $K \subset 4D$ then

$$N(K, D) \leq N(D, \frac{1}{16}K^\circ). \tag{1.4.8}$$

(2) If $D \subset 4K$ then

$$N(D, K) \leq N(K^\circ, \frac{1}{16}D). \tag{1.4.9}$$

Proof Notice that for any symmetric convex body T one has $T \cap T^\circ \subset D$, and similarly $D \subset T + T^\circ$. For (1), write, by Lemma 1.3.5 and the above, that

$$N(K, D) \leq N(K, 2K \cap \frac{1}{2}K^\circ) \leq N(K, \frac{1}{4}K^\circ). \tag{1.4.10}$$

(In fact we could put 2 instead of 4 since the centers of the covering can be assumed to lie inside K, and the same reasoning as in Fact 1.3.6 applies.) In the case $K \subset 4D$ this implies that

$$N(K, D) \leq N(D, \frac{1}{16}K^\circ).$$

For (2) write, by Lemma 1.3.4 and the above, that

$$N(D, K) \leq N(\frac{1}{2}K + 2K^\circ, K) \leq N(2K^\circ, \frac{1}{2}K)$$

In the case $D \subset 4K$ this implies that

$$N(D, K) \leq N(K^\circ, \frac{1}{16}D).$$

$\square$

To complete the proof of Theorem 1.2.1, combine the Lemma 1.4.4 with the reduction Proposition 1.4.1. The constants we arrive at are $\alpha = c' = \frac{1}{16\times 8} = \frac{1}{128}$ and $\beta = C' = 10$. $\square$

1.4.3 An alternative method of proof

Our original proof (appearing in the paper [AMS3]) of Theorem 1.2.1 was somewhat different. Although the proof described above is simpler, it has the disadvantage that it yields the universal constant 10 in the power, which is not optimal. The original proof was more complicated, but ensured any constant $2+\varepsilon$ in the exponent (with the other constant depending on ε). In the original proof one makes much more use of the euclidean structure. We devote this section to an exposition of this second method of proof. There is more to this other proof than merely an improvement of the constant. On the way we will encounter an open conjecture (posed by Milman and Szarek in [MSz1]); if this conjecture were verified, this second method would lead to a constant $(1+\varepsilon)$ in the exponent for any ε. Moreover, we feel that the importance of this conjecture (misleadingly named "the Geometric Lemma"; it is for the time being unverified) transcends its relevance to entropy numbers. Other parts of this alternative proof are also of independent interest, and will reappear in Chapter 3 of this thesis.

This second proof of the theorem consists of three parts, and then a fourth step to get the constant $\beta = 2+\varepsilon$ on both sides of inequality (1.2.1) in Theorem 1.2.1. The first part is based on a fact which has been formulated in our paper [AMS1], in which we establish duality up to some factor γ depending on the body K. Next, this step is iterated, each time applied to a different body (for example, a multiple of K intersected with a euclidean ball of some radius), and - similarly to the first proof - we bound the covering number by a product of covering numbers of polars. In the third and last step we collapse this product to a product of only a constant number of factors, establishing duality with absolute constants. Since this is a two sided inequality, almost every statement is divided into two parts. However, there is generally no interplay between the two arguments, and the proofs of the two sides of the inequality can be read independently.

1.4.3.1 The first step (in the second method of proof)

For a symmetric convex body $K \subset \mathbb{R}^n$, denote by k the logarithm of its covering number (so that $N(K, D) = 2^k$), and define the parameter

$$\gamma(K) := \max\{1, M^*(K \cap D)\sqrt{\frac{n}{k}}\},$$

where $M^*(A)$ denotes half the mean width of the set A, that is, $M^*(A) = \int_{S^{n-1}} \sup_{y\in A}\langle u, y\rangle \, d\mu(u)$, with μ the normalized Haar measure on the sphere S^{n-1}.

The first step of this proof of Theorem 1.2.1 is a duality result involving the parameter γ instead of a universal constant α. The following lemma is a combination of two statements, the first of which appeared in [AMS1] and the second in [MSz1].

Lemma 1.4.5 (First step) *There exists a universal constant $c_2 > 0$ and, for every $\varepsilon > 0$, a constant $C_2(\varepsilon) > 0$ such that, for any dimension n and for any symmetric convex body $K \subset \mathbb{R}^n$, denoting $\gamma = \gamma(K)$, we have*

$$N(K, D) \leq N(D, \frac{c_2}{\gamma} K^\circ)^3 \tag{1.4.11}$$

and

$$N(D, C_2(\varepsilon)\gamma K^\circ) \leq N(K, D)^{1+\varepsilon}. \tag{1.4.12}$$

To prove this lemma we need the following well known result by König and Milman [KöM]; a new proof of this type of result is described in Section 1.5.5.3.

Lemma 1.4.6 *There exists a universal constant C' such that for any dimension n, and any two symmetric convex bodies K and T in $\mathbb{R}^n$, one has*

$$\frac{1}{C'} \leq \left(\frac{N(K, T)}{N(T^\circ, K^\circ)}\right)^{\frac{1}{n}} \leq C', \tag{1.4.13}$$

(In fact the bodies need not be symmetric; it suffices to require that their centroids are 0, see [MPa].)

Proof of Lemma 1.4.5 We include only the proof of (1.4.11). The proof of the second inequality (1.4.12) appeared in [MSz1], with 2 instead of $(1 + \varepsilon)$ in the exponent. The same proof, with minor changes, gives $(1 + \varepsilon)$. To prove (1.4.11), denote $N := N(K, D)$. Our goal is to prove that $N(D, c_2\gamma^{-1}K^\circ) \geq N^{1/3}$. Let $C_1 = (c_2)^{-1}/2$ (our choice of c_2 will be specified later). Assume first that $N(K, C_1\gamma D) \leq N^{2/3}$. A result of Tomczak-Jaegermann [Tom] states that $N(K, D) \leq N(K, \theta D)N(D, \frac{1}{2\theta}K^\circ)$ for all $\theta > 0$. To prove it, combine inequality (1.4.10), with a factor 2 instead of 4 as explained in the remark below it, with the inequality $N(K, 1/2K^\circ) \leq N(K, \theta D)N(\theta D, \frac{1}{2}K^\circ)$. This implies then that

$$N(D, (2C_1\gamma)^{-1}K^\circ) \geq N^{1/3},$$

which yields (1.4.11).

It remains to handle the case $N(K, C_1\gamma D) > N^{2/3}$. Fact 1.3.7 implies that in this case there exists a $C_1\gamma$-separated (in the euclidean metric) subset S of K of cardinality $> N^{2/3}$. We set $k_1 := \lfloor c_1 \log N \rfloor$, with $c_1 = (3 \log C')^{-1}$, where C' comes from (1.4.13). [This will ensure that when we use (1.4.13) in dimension k_1, we will get a meaningful result; notice also that we may assume that $k_1 < n$, as otherwise $\log N \geq {c_1}^{-1}n$ and the answer follows easily from

(1.4.13).] Now consider a random orthogonal projection P_E of K onto a k_1-dimensional subspace E, distributed uniformly with respect to the Haar measure on the corresponding Grassmanian. It is well known that under such projections a distance between two points is typically multiplied by a factor close to $\sqrt{k_1/n}$. Moreover, from the results described in [Art1] (and in Section 3.2.6 of this thesis) which are an isomorphic extension of the Johnson-Lindenstrauss Lemma [JoLi], we know that if $S \subset \mathbb{R}^n$ is a finite set, then – with probability close to 1 – *all* the distances between elements of S are multiplied by factors $\geq c_2\sqrt{k_1/n}$, where $c_2 > 0$ depends only on (an upper bound for) the ratio $\log|S|/k_1$; in particular under our assumptions $c_2 \in (0, 1]$ is universal. It follows that, with probability close to 1, the image $P_E S$ of the $C_1\gamma$-separated set S is $c_2 C_1\sqrt{k_1/n}\,\gamma$-separated. Noting that $P_E S \subset P_E K \subset E$, we deduce that

$$N(P_E K, C_2\sqrt{k_1/n}\gamma D_E) > N^{2/3},$$

where D_E denotes the euclidean unit ball in E and $C_2 = c_2 C_1 = c2c^{-1}/2$. However, we are now in a position to use (1.4.13) in dimension k_1, which (by our choice of k_1) implies

$$N(D_E, C_2\sqrt{k_1/n}\gamma K^\circ \cap E) > N^{1/3}. \tag{1.4.14}$$

All that is left is to return to the original dimension. For this we use a Dvoretzky theorem-type assertion, which claims that a random projection of $K \cap D$ onto dimension k_1 is, with probability close to 1, contained in the euclidean ball of radius $C_3 \max\{M^*(K \cap D), \sqrt{k_1/n}\}$, where $C_3 \geq 1$ is a universal constant (for reference see [Mi2] or Fact 3.1 in [MSz1]). Dualizing this (and using the definitions of γ and k_1) we obtain

$$C_4\sqrt{k_1/n}\;\gamma\;\mathrm{conv}(K^\circ \cup D) \cap E \supset D_E$$

(with $C_4 = C_3 \max\{1, \sqrt{2/c_1}\}$), which combined with (1.4.14) yields

$$N(C_4\mathrm{conv}(K^\circ \cup D) \cap E, C_2 K^\circ \cap E) > N^{1/3}.$$

We now choose the original c in the statement of the theorem: $c = (C_4)^{-1}c_2/8$. This means precisely that $C_2 = 4C_4$. Then

$$\begin{aligned} N^{1/3} &< N(\mathrm{conv}(K^\circ \cup D) \cap E, 4K^\circ \cap E) \\ &\leq N(\mathrm{conv}(K^\circ \cup D) \cap E, 2K^\circ) \\ &\leq N(K^\circ + D, 2K^\circ) \\ &\leq N(D, K^\circ), \end{aligned}$$

where for the second inequality we use the fact that, for every $x \in \mathbb{R}^n$, the set $(x + 2K^\circ) \cap E$ is contained in a translate of $4K^\circ \cap E$. The proof is now complete; notice that in the second case the parameter γ does not enter into the final formulae and we obtain duality with absolute constants. This completes the proof of (1.4.11). $\square$

For a general body $K \subset \mathbb{R}^n$ the parameter γ can be as large as $\sqrt{\frac{n}{k}}$. In Observation 1.4.7 below we explain why we can restrict our considerations to a certain special class of convex bodies, namely the convex hulls of not too many points. In the rest of this sub-section we will show that for bodies in this class we can find good bounds for γ.

Observation 1.4.7 *For any convex body K, any set $S \subset K \subset S + D$ of cardinality $N(K, D)$ and any $\rho > 0$ we have, denoting $T = \mathrm{conv}(S)$,*

$$N(D, (2\rho + 2)K^\circ) \le N(D, \rho T^\circ).$$

Similarly, if $N(K, D) > 1$, then we can find $S \subset K$ of cardinality $N(K, D)$ such that $\mathrm{diam}(S) = \mathrm{diam}(K)$ *and that, denoting $T = \mathrm{conv}(S)$, we have*

$$N(K, D) \le N(T, \frac{1}{2}D).$$

Remark: The argument does not require that the original body lie in a finite dimensional space (whereas a convex hull of finitely many points obviously does). In particular, this shows the equivalence of the operator theoretic formulation of the duality conjecture and the finite dimensional analogue with universal constants.

Proof Obviously $T \subset K \subset T + D$. Denoting $N(D, \rho T^\circ) = N$, we can pick a ρT°-net $\{y_i\}_{i=1}^N$ for D, i.e., $D \subset \cup_{i=1}^N y_i + \rho T^\circ$. We want to pass to a net *inside* the ball D. For this notice that $y_i + \rho T^\circ$ intersects D, say at a point z_i, and that $\{z_i\}_{i=1}^N$ is a $2\rho T^\circ$-net for D. We claim that $\{z_i\}_{i=1}^N$ is a $(2\rho + 2)K^\circ$-net for D. Indeed, for every $y \in D$ there exists a z_i in the net such that $y - z_i \in 2\rho T^\circ$, i.e., $\sup_{x \in T}(y - z_i, x) \le 2\rho$. Hence (using that $y - z_i$ is in $2D$) we have $\sup_{x \in T+D}(y - z_i, x) \le 2\rho + 2$, which means precisely that $\|y - z_i\|_{(T+D)^\circ} \le 2\rho + 2$. In particular, since $K \subset T + D$, we see that $\|y - y_i\|_{K^\circ} \le \rho + 2$, as required. We conclude that $N(D, (2\rho + 2)K^\circ) \le N$, and this verifies the first part of the observation.

For the second part, we denote this time $N = N(K, D)$ and pick a 1-separated set $\{x_i\}_{i=1}^N$ in K which realizes the diameter. We do this simply by choosing two points, the distance between which is the diameter of K, and completing them to a 1-separated set of cardinality N. Again, this is possible since a maximal separated set has at least as many elements as the minimal covering. Denote $T = \text{conv}\{x_i\}$. Since $\{x_i\}$ were 1-separated, $N(T, \frac{1}{2}D) \geq N$. This completes the demonstration of the observation. $\square$

The following proposition is an estimate for $\gamma(K)$ which is valid whenever K is the convex hull of $\leq 2^k$ points in RD and, in addition, has a covering number $\leq 2^k$. The general conjecture, which still remains open, is that for this class of bodies the parameter γ is bounded by a universal constant, regardless of the diameter of the body. If this were true, Lemma 1.4.5 and Observation 1.4.7 would imply the duality of entropy numbers (with $1+\varepsilon$ in the exponent). We discuss the conjecture in subsection 1.4.3.4; for a more elaborate discussion and related results we refer the reader to [MSz1].

Proposition 1.4.8 (An $O(\log^3 R)$ estimate for γ) *There exists a universal constant C_0 such that if a set $S \subset RD \subset \mathbb{R}^n$ (for some $R > 1$) consists of 2^k points, and if $N(K, D) \leq 2^k$ for $K = \text{conv}S$, then*

$$M^*(K \cap D) \leq C_0 (\log R)^3 \sqrt{\frac{k}{n}}.$$

We provide only a sketch of the proof of the proposition whose full proof was established in [MSz1].

Proof of Proposition 1.4.8 Formally the proof is divided to four consequent steps:
Step 1. If for $\varphi : R^+ \to R^+$ the conditions $\text{conv}_{i \leq 2^k}\{x_i\} = K \subset RD \subset R^n$, $N(K, D) \leq 2^k$ and $k \leq \varphi(R)n$ imply that $\frac{1}{2}D \not\subset K$, then the conditions $\text{conv}_{i \leq 2^k}\{x_i\} = K \subset RD \subset R^n$ and $N(K, D) \leq 2^k$ imply $M^*(K \cap D) \leq C(\varphi(R))^{-1/2}\sqrt{\frac{k}{n}}$.
Step 2. For $\alpha D \subset \text{conv}_{i \leq 2^{k_1}}\{x_i\} = K \subset RD \subset R^n$ with $N(K, D) \leq 2^k$ there exists $K_1 \subset R^{n_1}$ for $n_1 \geq n/2$ with $\alpha D_{n_1} \subset \text{conv}_{i \leq 2^{k_1}}\{y_i\} = K_1 \subset RD_{n_1}$, $N(K_1, D_{n_1}) \leq 2^k$ and $M^*(K_1) \leq C(1 + \log(R/\alpha))$.
Step 3. If $\text{conv}_{i \leq 2^{k_1}}\{x_i\} = K \subset RD \subset R^n$ and $N(K, D) \leq 2^k$, then $M^*(K \cap D) \leq C(R\sqrt{\frac{k}{n}}\sqrt{\frac{k_1}{n}})^{1/2}$.
Step 4. If $\text{conv}_{i \leq 2^k}\{x_i\} = K \subset RD \subset R^n$, $N(K, D) \leq 2^k$ and $k \leq [\frac{c}{\log R}]^6 n$, then $\frac{1}{2}D \not\subset K$.
It is clear that Steps 1 and 4 together complete the proof of Proposition 1.4.8. We will not give the proofs of all four steps; we only show how to track down the actual points which are being used. We first take $\text{conv}_{i \leq 2^k}\{x_i\} = K \subset RD$ in R^n with $N(K, D) \leq 2^k$. We let $n_1 = cM^*(K \cap D)^2 n$ and project K randomly

to an n_1-dimensional subspace. We denote $\frac{1+\varepsilon}{M^*(K\cap D)}P_{n_1}K = K_1 \subset R^{n_1}$. Then $\frac{1}{2}D \subset K_1 \subset RD$ and $N(K_1, D) \le 2^k$. We now pick a specific subspace of R^{n_1} of dimension $n_2 \ge n_1/2$ in the following way: let $u \in GL_{n_1}$ be so that MM^* of uK_1 is minimal. Then pick a subspace for which $P_{n_2}(uD_{n_1}) = \lambda D_{n_2}$. (This is always possible for ellipsoids.) Denote now $K_2 = \lambda^{-1}P_{n_2}uK_1$. Then K_2 is the convex hull of 2^k points, $N(K_2, D) \le 2^k$, $\frac{1}{2}D \subset K_2 \subset RD$ and $M^*(K_2) \le C(1+\log R)$. For this K_2 one can show that it must have $M^*(K_2 \cap D) \le (\log R)^{3/2}(\frac{k}{n_2})^{1/4}$ (in this part K_2 is decomposed in a sophisticated manner using a theorem of Talagrand). However, it contains $\frac{1}{2}D$, and so this gives us a condition $k \ge \varphi(R)n_2$, which, when you write what n_2 is, translates to precisely $M^*(K\cap D) \le \sqrt{\frac{c}{\varphi(R)}}\sqrt{\frac{k}{n}}$. □

Lemma 1.4.5, Observation 1.4.7 and the above Proposition 1.4.8 can be combined as follows. Denote $\psi(x) = 2C_2(C_0 \log^3 x + 1) + 2$, where $C_2 = C_2(1)$ comes from Lemma 1.4.5.

Corollary 1.4.9 (Duality up to $\psi(R)$) *Let K be a symmetric convex body in $\mathbb{R}^n$ and assume that for some $R > 1$ we have $K \subset RD \subset \mathbb{R}^n$. Then*

$$N(D, \psi(R)K^\circ) \le N(K, D)^2 \tag{1.4.15}$$

and

$$N(K, D) \le N(D, (1/\psi(R))K^\circ)^3. \tag{1.4.16}$$

1.4.3.2 Second Step - An iterating scheme

In this section we present an iterating procedure that gives a bound for the covering number. The first lemma is based on a simple geometric iteration procedure (and admits a variant which is valid in the non-euclidean case; Lemma 1.4.2 above).

Lemma 1.4.10 (Iterating procedure) *For a symmetric convex body $K \subset \mathbb{R}^n$ and any sequence $R_0 < R_1 < \cdots < R_s$,*

$$N(D, R_0K^\circ) \le N(D, R_sK^\circ)\prod_{j=0}^{s-1} N(D, \frac{R_j}{2}(K\cap R_{j+1}D)^\circ), \tag{1.4.17}$$

and

$$N(K, R_0D) \le N(K, R_sD)\prod_{j=0}^{s-1} N(2K\cap R_{j+1}D, R_jD). \tag{1.4.18}$$

Proof The proof is almost identical to the proof of Lemma 1.4.2. For (1.4.17) consider the following inequality

$$N(D, R_0 K^\circ) \le N(D, \frac{R_0}{2}\mathrm{conv}(K^\circ \cup \frac{1}{R_1}D))N(\frac{R_0}{2}\mathrm{conv}(K^\circ \cup \frac{1}{R_1}D), R_0 K^\circ),$$

which follows from Fact 1.3.3. Rewriting the first term on the right hand side, changing the convex hull in the second term to the Minkowski sum of the sets (which is bigger and thus harder to cover) and using Lemma 1.3.4 leads to

$$N(D, R_0 K^\circ) \le N(D, \frac{R_0}{2}(K \cap R_1 D)^\circ)N(D, R_1 K^\circ).$$

Repeating the above argument $(s-1)$ times more yields (1.4.17).

To show (1.4.18) first notice that

$$N(K, R_0 D) \le N(K, R_1 D)N(R_1 D \cap 2K, R_0 D),$$

where we have used Lemma 1.3.6 and Lemma 1.3.3. Iterating this inequality gives (1.4.18). The proof of Lemma 1.4.10 is thus complete. □

We now choose the sequence (R_j). In fact, we will choose two different sequences, each corresponding to a different inequality in the main theorem. There is much freedom in this choice, and we do not suggest that the choice below is optimal in any sense.

For the first sequence, let R_0 be a large constant to be specified later. Define R_{j+1} by the formula

$$\frac{\sqrt{R_j}}{2} = \psi\left(\frac{R_{j+1}}{\sqrt{R_j}}\right).$$

Remembering that $\psi(x) = 2C_2(C_0(\log x)^3 + 1) + 2$, the above means that

$$R_{j+1} = \sqrt{R_j}\ \exp\left(((\sqrt{R_j} - 4 - 4C_2)/4C_2C_0)^{1/3}\right).$$

In particular, if R_0 is large enough then this sequence increases to ∞. (This is needed since we will later use the fact that $N(D, R_j K^\circ) = 1$ for j large enough.) Corollary 1.4.9 together with Lemma 1.4.10 implies the following

Corollary 1.4.11 *With the above choice of the sequence (R_j) we have, for every symmetric convex body K,*

$$N(D, R_0 K^\circ) \le N(D, R_s K^\circ)\prod_{j=0}^{s-1} N(K \cap R_{j+1}D, \sqrt{R_j}D)^2. \tag{1.4.19}$$

Proof To deduce Corollary 1.4.11 from Lemma 1.4.10 we only need to explain the inequality

$$N(D, \frac{R_j}{2}(K \cap R_{j+1}D)^\circ) \le N(K \cap R_{j+1}D, \sqrt{R_j}D)^2.$$

To this end, rewrite

$$\begin{aligned} N\left(D, \frac{R_j}{2}(K \cap R_{j+1}D)^\circ\right) &= N\left(D, \frac{\sqrt{R_j}}{2}\left(\frac{K}{\sqrt{R_j}} \cap \frac{R_{j+1}}{\sqrt{R_j}}D\right)^\circ\right) \\ &= N\left(D, \psi\left(\frac{R_{j+1}}{\sqrt{R_j}}\right)\left(\frac{K}{\sqrt{R_j}} \cap \frac{R_{j+1}}{\sqrt{R_j}}D\right)^\circ\right) \\ &\le N\left(\frac{K}{\sqrt{R_j}} \cap \frac{R_{j+1}}{\sqrt{R_j}}D, D\right)^2 \\ &= N(K \cap R_{j+1}D, \sqrt{R_j}D)^2, \end{aligned}$$

where for the inequality we used (1.4.15) of Corollary 1.4.9. □

For the proof of the other inequality in Theorem 1.2.1 we have a different condition on the sequence to make this type of argument work. Again, let R'_0 be a big constant to be specified later. Define R'_{j+1} by

$$\psi\left(\frac{R'_{j+1}}{R'_j}\right) = \frac{\sqrt{R'_j}}{2},$$

which can be rewritten as $R'_{j+1} = R'_j \exp\left(\left(\left(\sqrt{R'_j} - 4 - 4C_2\right)/4C_2C_0\right)^{1/3}\right)$. Again, it is clear that this sequence is increasing to ∞.

Corollary 1.4.12 *With the above choice of a sequence R'_j we have, for every convex symmetric body K,*

$$N(K, R'_0D) \le N(K, R'_sD) \prod_{j=0}^{s-1} N(D, \sqrt{R'_j}(K \cap \frac{R'_{j+1}}{2}D)^\circ)^3. \tag{1.4.20}$$

Proof Again, we will use Lemma 1.4.10 together with Corollary 1.4.9. Here we should explain the inequality

$$N(2K \cap R'_{j+1}D, R'_jD) \le N(D, \sqrt{R'_j}(K \cap \frac{R'_{j+1}}{2}D)^\circ)^3.$$

This is even simpler since

$$\begin{aligned}
N(2K \cap R'_{j+1}D, R'_jD) &= N\left(\frac{2}{R'_j}K \cap \frac{R'_{j+1}}{R'_j}D, D\right) \\
&\leq N\left(D, \frac{R'_j}{2\psi\left(\frac{R'_{j+1}}{R'_j}\right)}(K \cap \frac{R'_{j+1}}{2}D)^\circ\right)^3 \\
&= N(D, \sqrt{R'_j}(K \cap \frac{R'_{j+1}}{2}D)^\circ)^3,
\end{aligned}$$

where for the inequality we used (1.4.16) of Corollary 1.4.9. □

1.4.3.3 Telescoping

In this last step we collapse the long products of covering numbers appearing in (1.4.19) and (1.4.20) to products consisting of just two terms. The largest R_s (respectively R'_s) will be chosen to exceed the diameter of the set, and so the terms $N(K, R'_sD)$ and $N(D, R_sK^\circ)$ will both equal 1. We use, as in the first proof, the two super-multiplicativity inequalities for covering numbers in Lemma 1.4.3, this time only for the body $T = D$.

Proof of the Main Theorem We will successively apply Lemma 1.4.3 to the long products in (1.4.19) and (1.4.20). However, as in the proof of Proposition 1.4.1, an additional trick is required since for two neighboring factors in the products the condition of Lemma 1.4.3 does not hold, and so they cannot be "collapsed." For example, for two such factors in (1.4.19) one has $a = \sqrt{R_j}$ and $B = R_j$, and so one cannot hope for $a > 3B$. The trick is to split the product into two parts, by grouping separately the factors corresponding to the odd and the even j's (thus taking the 'l' appearing in the proof of Proposition 1.4.1 to be equal 2 suffices). The growth of R_j is fast enough so that the conditions of Lemma 1.4.3 are satisfied for each two consecutive odd factors, and for each two consecutive even factors. We provide details for the product from (1.4.19); the analysis of (1.4.20) is fully analogous.

First choose s to be the smallest even number so that $R_s > \mathrm{diam}(K)$. Then the product in (1.4.19) which bounds $N(D, R_0K^\circ)$ can be written as (we omit the power 2 for the moment)

$$\prod_{j=1}^{s/2} N(K \cap R_{2j}D, \sqrt{R_{2j-1}}D) \prod_{j=1}^{s/2} N(K \cap R_{2j-1}D, \sqrt{R_{2j-2}}D).$$

For the first collapsing step in each of the two sub-products we need to check that $\sqrt{R_{s-1}} > 3R_{s-2}$, and that $\sqrt{R_{s-2}} > 3R_{s-3}$.

When $R_{j+1} = \sqrt{R_j}\exp\left(((\sqrt{R_j}-4-4C_2)/4C_2C_0)^{1/3}\right)$, these conditions clearly hold as long as R_{s-3} is larger than some numerical constant C. Since $R_j > R_0$ for $j > 0$, it is enough to start with R_0 which is big enough. To be able to later compare the obtained expressions with $N(K, D)$, we also insist that $R_0 \geq 16$. After this first step, using Lemma 1.4.3, the product becomes (bounded by)

$$N(K \cap R_s D, \frac{\sqrt{R_{s-3}}}{4} D) \prod_{j=1}^{s/2-2} N(K \cap R_{2j} D, \sqrt{R_{2j-1}} D)$$
$$N(K \cap R_{s-1} D, \frac{\sqrt{R_{s-4}}}{4} D) \prod_{j=1}^{s/2-2} N(K \cap R_{2j-1} D, \sqrt{R_{2j-2}} D).$$

From here onward all the steps are the same; we just need to make sure at each stage j that

$$(\sqrt{R_j}/4)/(R_{j-1}/2) \geq 3. \tag{1.4.21}$$

As before, this is indeed satisfied if $R_j > C$, which is assured since we insist that $R_0 > C$. We point out, again, that the factors $1/4$ in $b/4$ do not accumulate, but enter into the quantity a of the next step. Continuing this way with all the factors of these products, we arrive at

$$N(D, R_0 K^\circ) \leq N(K \cap R_s D, \frac{\sqrt{R_0}}{4} D)^2 N(K \cap R_{s-1} D, \frac{\sqrt{R_1}}{4} D)^2,$$

(we have inserted back the power 2 in (1.4.19)) which, having insisted that $R_0 \geq 16$, implies

$$N(D, R_0 K^\circ) \leq N(K, D)^4.$$

Similarly, in the other direction we use Corollary 1.4.12 and inequality (1.4.4) to obtain

$$N(K, R_0 D) \leq N(D, K^\circ)^6,$$

and the proof of the main theorem is complete. $\square$

1.4.3.4 To get the constant $2 + \varepsilon$

In this section we explain how to improve the constant $\beta = 6$ in (1.2.1) that we have obtained in the previous section to a constant $\beta = 2(1 + \varepsilon)$. The proof presented above is somewhat non-symmetric. As described, in one of the inequalities we get $\beta = 6$, and in the other $\beta = 4$. This $\beta = 4$ can be improved to $2(1 + \varepsilon)$ if we continue working with a general ε in Lemma 1.4.5 and do not specify (as we did only to lighten the notation) $\varepsilon = 1$. However, as stated above, we cannot in a straightforward way obtain $\beta = 2(1+\varepsilon)$ in the other inequality. Below we explain how to arrive at $\beta = 2(1+\varepsilon)$ there as well. We take this opportunity to elaborate upon the conjecture called "the Geometric Lemma," which we mentioned in passing in Section 1.4.3.1.

Conjecture 1.4.13 (Geometric Lemma) *There exists an absolute constant C_0 such that for every dimension n and for every set $S = \{x_i\}_{i=1}^{2^k} \subset \mathbb{R}^n$ verifying $N(K, D) \le 2^k$, where $K = \mathrm{conv}(S)$, we have*

$$M^*(K \cap D) \le C_0 \sqrt{\frac{k}{n}}.$$

Notice that this is precisely a version of Proposition 1.4.8 with no dependence on the radius of the set. This conjecture is, for the time being, an open question. If it is true, it implies that Theorem 1.2.1 is true with $\beta = 1 + \varepsilon$ for any ε.

Considering now the dual situation, we can formulate a dual version of the Geometric Lemma, substituting the condition $N(K, D) \le 2^k$ by the condition $N(D, K^\circ) \le 2^k$. However, having proved the duality of entropy numbers, it is easy to see that the Geometric Lemma and its dual version are formally equivalent. Moreover, every estimate such as Proposition 1.4.8 can be applied now to the dual situation. For example the main theorem together with Proposition 1.4.8 gives the following

Proposition 1.4.14 (A dual $O((\log R)^3)$ estimate for $M^*(K \cap D)$) *There exists a universal constant C_0 such that if a set $S \subset RD \subset \mathbb{R}^n$ consists of 2^k points, and if $N(D, K^\circ) \le 2^k$ for $K = \mathrm{conv}S$, then*

$$M^*(K \cap D) \le C_0 (\log R)^3 \sqrt{\frac{k}{n}}.$$

Proof Indeed, Assume that $N(D, K^\circ) \le 2^k$, this implies that $N(\alpha K, D) \le 2^{\beta k}$ (these are $\alpha < 1$ and $\beta > 1$ from Theorem 1.2.1). We now use Proposition 1.4.8, together with the assumption that $\alpha S \subset \alpha R D$ and consists of 2^k points (which is less than $2^{k\beta}$). Thus Proposition 1.4.8 implies that

$$M^*(\alpha K \cap D) \le C_0 (\log(\alpha R))^3 \sqrt{\frac{\beta k}{n}}.$$

This in turn implies

$$M^*(K \cap D) \le C_0 \frac{\sqrt{\beta}}{\alpha} (\log R)^3 \sqrt{\frac{k}{n}}.$$

$\square$

We can thus define a new parameter, $\gamma'(K)$, to be

$$\gamma'(K) := \max\{1, M^*(K \cap D)\sqrt{\frac{n}{\log N(D, K^\circ)}}\},$$

and obtain the following

Lemma 1.4.15 (Dual First step) *For every $\varepsilon > 0$ there is a constant $C_2' = C_2'(\varepsilon)$ such that for any dimension n and for any symmetric convex body $K \subset \mathbb{R}^n$, denoting $\gamma' = \gamma'(K)$ we have*

$$N(K, C_2'\gamma' D) \leq N(D, K^\circ)^{1+\varepsilon}. \tag{1.4.22}$$

Proof We need to repeat the argument of Lemma 1.4.5. Assume that $N(D, K^\circ) \leq N$, then there is a set S of cardinality N with $D \subset S + K^\circ$. Intersecting with any k-dimensional subspace E we have

$$E \cap (D + K^\circ) \subset (S + 2K^\circ) \cap E.$$

We choose $k = c \log N$ with $c = \varepsilon \log C'$, where C' is from Lemma 1.4.6. Similarly to the proof of Lemma 1.4.5, we have for a random E of dimension k that $D_E \subset C\gamma'\sqrt{\frac{k}{n}}(K^\circ + D) \cap E$, where C depends only on c, that is, only on ε and C'. Therefore $N(D_E, 2C\sqrt{\frac{k}{n}}\gamma' K^\circ \cap E) \leq N$. Together with Lemma 1.4.6 this implies that

$$N(P_E K, 2C\sqrt{\frac{k}{n}}\gamma' D_E) \leq N(C')^k = N^{1+\varepsilon}.$$

Now we just have to notice that, by a Johnson-Lindenstrauss type argument, if for some C_1 we know that $N(K, C_1 D) \geq N^{1+\varepsilon}$ then for any $k \geq c \log N$ and a random projection P_k, one has $N(P_k K, c' C_1 \sqrt{\frac{k}{n}} D) \geq N^{1+\varepsilon}$, where c' depends only on ε and c. This means that any $C_1 \geq 2C\gamma'/c'$ cannot work, or, in other words, that $N(K, 2C\gamma'/c' D) \leq N^{1+\varepsilon}$. The proof of the lemma is complete. □

Employing the same line of argument as earlier, but using Proposition 1.4.14 as an estimate on γ', and inequality (1.4.22) at every step, we obtain $\beta = 2 + \varepsilon$, instead of $\beta = 6$, also in the other inequality involved in the duality of metric entropy. The details are omitted.

1.5 Generalizations

In this section we introduce a new notion of separation, which we call convex separation or convexified packing. We prove a duality theorem for the cardinality of maximal convexly separated sets. We then show that in a special case the new notion of maximal convexly separated sets and the usual covering numbers are equivalent. We use this to prove Conjecture 1.2.2 (the duality conjecture) for the case in which either the domain or the range of the operator has some non trivial type. This result is formulated in Theorem 1.5.5 below. It generalizes Theorem 1.2.1, described in sections 1.2 and 1.4, which verified the conjecture when either the domain or the range of the operator was a Hilbert space. The case in which one of the two spaces has some non trivial type (definitions below) is quite general, and includes the case of uniformly convex or uniformly smooth bodies, so, in particular, it includes the case of all l_p or L_p spaces for $1 < p < \infty$. To achieve this we have joined the approach from [AMS3] with an older approach to this problem which originated in the paper [Tom] and was generalized in [BPST].

What we find to be the importance of this method is that it splits the duality theorem into two parts. First, we have duality for convex separation, which is simply a consequence of the Hahn-Banach theorem. Secondly, there are arguments connecting convex separation and coverings, and iteration procedures, but all these occur each time in a *given space*. We feel that whereas until now connections between covering numbers and covering numbers of duals were surprising and even somewhat mystical facts, this new method enables us to extract the (one!) step where a duality connection between the geometry of a space and its dual appears, and understand it. The rest of the proofs are geometrical considerations; while often delicate and involved, they are always set in a given normed space, and thus are conceptually simpler.

The notion of type (and of cotype) in the context of Banach spaces goes back to Nordlander, and arose also in the work of Stein concerning inequalities for maximal operators. It was formally introduced by Hoffmann-Jørgensen in [HJ]. The direction was continued by B. Maurey and the the theory was built by Maurey and Pisier, see for example [Mau] and [MaPi]. Later in the 70's and 80's the study was continued by many authors.

The definition is as follows. A Banach space X is said to have type $1 \le p \le 2$ if there exists a constant C such that for all n and all $x_1, \dots, x_n$ in X one has

$$|| \sum \varepsilon_i x_i ||_{L_2(X)} \le C (\sum ||x_i||^p)^{1/p}, \tag{1.5.1}$$

where ε_i are i.i.d. random variables with values ± 1 with probabilities $1/2$. It can be easily checked that every space has type 1, and so if a space has some larger type $p > 1$ we say that it has non-trivial type. The best constant C in inequality (1.5.1) is called the type p constant of X, and is denoted by T_p (or $T_p(X)$). We say that a convex body K has type p with constant T_p if this is true for the normed space X whose unit ball is K.

The section is organized as follows. In Section 1.5.1 we define the notion of convex separation. In Section 1.5.2 we show duality for the cardinality of a maximal convexly separated set. In Section 1.5.3 we use Maurey's Lemma to show that in a special case this cardinality and the covering number are equivalent up to universal constants. In Section 1.5.4 we complete the proof of the generalization. In section 1.5.5 we explain how other results concerning covering numbers can be obtained from the point of view of convex separation, in particular a result similar to Lemma 1.4.6. By this we want to emphasize that essentially all known facts about duality of covering numbers can be obtained from the point of view of convex separation, and Proposition 1.5.1 of Section 1.5.2 is the true underlying reason for the duality phenomenon.

1.5.1 Defining Convex Separation

A notion very similar to covering numbers is that of the cardinality of a maximal separated set, which we recalled in Section 1.2 and denoted by $M(K,T)$. (It is the maximal number of points in K such that the distance between each two, in the norm whose unit ball is T, is greater than 1). Since (see Fact 1.3.7) $M(K,2T) \leq N(K,T) \leq M(K,T)$, and since in the duality conjecture one permits a universal constant next to the body, one can reformulate the duality conjecture in terms of $M(K,T)$.

We now define a new number $\hat{M}(K,T)$, which can be thought of as the cardinality of a 'maximal convexly separated set', and plays the central role in this section:

$$\begin{aligned} \hat{M}(K,T) \quad = \quad & \max\{N : \exists x_1,\dots,x_N \in K \\ & \qquad \text{s.t. } (x_j+T)\cap \operatorname{conv}\{x_i, i<j\} = \phi\} \end{aligned} \tag{1.5.2}$$

We shall refer to any sequence satisfying the condition (1.5.2) as T-*convexly separated*. Leaving out the convex hull operation "conv" leads to the usual T-separated set. Thus we clearly have $\hat{M}(K,T) \leq M(K,T)$. We emphasize that, as opposed to the usual notions of packing and covering, the *order* of the points is important here.

The above definition is very natural from the point of view of complexity theory and optimization. A standard device in constructing geometric algorithms is a "separation oracle" (see e.g. [GLS]): if $x \notin K$, the algorithm returns a functional efficiently separating x from K. It is arguable that the quantity $\hat{M}(K,\cdot)$ correctly describes the complexity of a set K with respect to many such algorithms.

1.5.2 Duality for $\hat{M}(K,T)$

While it is still an open problem whether Conjecture 1.2.2 holds in full generality, the corresponding duality statement for convex separation is fairly straightforward. We have

Proposition 1.5.1 *For any pair of convex symmetric bodies $K, T \subset \mathbb{R}^n$ we have*

$$\hat{M}(K,T) \leq \hat{M}(T^\circ, K^\circ/2)^2.$$

We want to emphasize that this will turn out to be the only place where a connection between the geometry in the space and the geometry in the dual space needs to be established, and thus Proposition 1.5.1 is, perhaps, the heart of the duality phenomenon.

Proof of Proposition 1.5.1 Let $R := \sup\{\|x\|_T : x \in K\}$; i.e., R is the radius of K with respect to the gauge of T. We will show that
(i) $\hat{M}(T^\circ, K^\circ/2) \geq \hat{M}(K,T)/\lceil 4R \rceil$
(ii) $\hat{M}(T^\circ, K^\circ/2) \geq \lceil 4R \rceil$

Once the above are proved, Proposition 1.5.1 readily follows. To show (i), denote $N = \hat{M}(K,T)$ and let $x_1, \ldots, x_N$ be a T-convexly separated sequence in K. Then, by (the elementary version of) the Hahn-Banach theorem, there exist separating functionals $y_1, \ldots, y_N \in T^\circ$ such that

$$1 \leq i < j \leq N \Rightarrow \langle y_j, x_j - x_i \rangle = \langle y_j, x_j \rangle - \langle y_j, x_i \rangle \geq 1, \tag{1.5.3}$$

a condition which is in fact equivalent to (x_j) being T-convexly separated. Now $x_j \in K \subset RT$ and $y_j \in T^\circ$ imply that $-R \leq \langle y_j, x_j \rangle \leq R$, and hence dividing $[-R, R]$ into $\lceil 4R \rceil$ subintervals of length $\leq 1/2$ we may deduce that one of these subintervals contains $M \geq N/\lceil 4R \rceil$ of the numbers $\langle y_j, x_j \rangle$. To simplify the notation, assume that this occurs for $j = 1, \ldots, M$, that is

$$1 \leq i, j \leq M \Rightarrow -1/2 \leq \langle y_i, x_i \rangle - \langle y_j, x_j \rangle \leq 1/2. \tag{1.5.4}$$

Combining (1.5.3) and (1.5.4) we obtain for $1 \leq i < j \leq M$

$$\langle y_i - y_j, x_i \rangle = \langle y_i, x_i \rangle - \langle y_j, x_j \rangle + \langle y_j, x_j \rangle - \langle y_j, x_i \rangle \geq -1/2 + 1 = 1/2,$$

which is again a condition of type (1.5.3) and thus shows that the sequence $y_M, \ldots, y_1$ is $(K^\circ/2)$-convexly separated. Notice that the order has changed - indeed the above shows that y_i is separated from any convex combination $\sum_{j>i} \lambda_j y_j$ with $\sum_{j>i} \lambda_j = 1$ since

$$\|y_i - \sum_{j>i} \lambda_j y_j\|_{K^\circ} \geq \sum_{j>i} \lambda_j \langle y_i - y_j, x_i \rangle \geq 1/2.$$

Hence $\hat{M}(T^\circ, K^\circ/2) \geq M \geq N/\lceil 4R \rceil$, which is exactly the conclusion of (i).

To show (ii) we note that R is also the radius of T° with respect to the gauge of K°. Since we are in a finite-dimensional space, that radius is attained and so there is a segment $I := (-y, y) \subset T^\circ$ with $\|y\|_{K^\circ} = R$. This implies that $M(I, K^\circ/2) \geq \lceil 4R \rceil$. However, in dimension one separated and convexly separated sets coincide; this allows to conclude that $\hat{M}(T^\circ, K^\circ/2) \geq \hat{M}(I, K^\circ/2) \geq \lceil 4R \rceil$, as required. □

1.5.3 Connection to usual covering

The above Proposition 1.5.1 implies that if N and $\hat{M}$ are equivalent, we have duality also for usual covering numbers. However, it appears at the first sight that the notion of convex separation is much more restrictive than that of usual separation and, consequently, that – except for very special cases such as that of one-dimensional sets mentioned above – $\hat{M}(\cdot,\cdot)$ should be significantly smaller than $M(\cdot,\cdot)$ or $N(\cdot,\cdot)$. On the other hand, the reduction Proposition 1.4.1 implies that for the usual covering numbers duality problem, for two general convex bodies, it is enough to consider bodies K and T with $K \subset 4T$. Thus, it suffices to show that in *this* case the two kinds of covering numbers are equivalent.

In this special situation, of $K \subset 4T$, Maurey's Lemma gives a connection between N and $\hat{M}$ in the case where one of the two bodies has some non trivial type $p > 1$. We will show below that

Lemma 1.5.2 *If either K or T has a non trivial type $p > 1$ with type constant T_p, and in addition $K \subset 4T$, then, denoting $p^* = p/(p-1)$,*

$$\hat{M}(K, 2T) \leq N(K, T) \leq \hat{M}(K, \frac{1}{2}T)^{(16T_p)^{p^*}}.$$

To show this we use the well known result of B. Maurey (see e.g. [Pis4]), the proof of which we include below for the convenience of the reader.

Lemma 1.5.3 (Maurey's Lemma) *Let X be a space with type p, and unit ball T. Let m be an integer and assume that K is the convex hull of m points in RT. Then for any integer k*

$$N(K, 2k^{-1/p^*}T_p(X)RT) \leq m^k,$$

where p^* is the conjugate of p, that is $1/p + 1/p^* = 1$.

Proof It is well known, and can be easily proved by symmetrizing, that the fact that X has type p implies that for any k, and independent random variables

$Z_1, Z_2, \ldots, Z_k$ with values in X, one has

$$\mathbb{E}(||\sum_{i=1}^{k} Z_i - \mathbb{E}Z_i||) \leq 2T_p(X)(\sum_{i=1}^{k} \mathbb{E}||Z_i||^p)^{1/p}. \tag{1.5.5}$$

Denote $A = \{x_j\}_{j=1}^m$, and so the cardinality of A is m and $K = \mathrm{conv}(A)$. Let $x \in K$, then x is a convex combination of elements in A, say $x = \sum_{j=1}^m \lambda_j x_j$, with $\lambda_j \geq 0$ and $\sum \lambda_j = 1$. Define a random variable Z to assume the value x_j with probability λ_j, so that $\mathbb{E}Z = x$, and let $Z_1, \ldots, Z_k$ be independent copies of Z, where k is some large integer to be determined later. Equation (1.5.5) implies that

$$\mathbb{E}(||\sum_{i=1}^{k} Z_i - kx||) \leq 2T_p(X)k^{1/p}(\mathbb{E}||Z||^p)^{1/p} \leq 2T_p(X)k^{1/p}R.$$

Dividing by k, we see that there exists at least one k-tuple $x_{j_1}, \ldots, x_{j_k}$ in A^k with

$$||\frac{1}{k}\sum_{i=1}^{k} x_{j_i} - x|| \leq 2T_p(X)k^{-1/p^*}R.$$

Therefore, the set

$$\{\frac{1}{k}\sum_{i=1}^{k} x_{j_i} : \forall i \; x_{j_i} \in A\},$$

which is of cardinality at most m^k, is an εT-net for K where $\varepsilon = 2T_p(X)k^{-1/p^*}R$, as required. □

We will actually be using the following corollary

Corollary 1.5.4 *Let K or T have type p with type constant T_p, and assume that $K \subset 4T$. Then for any integer m and any $x_1, \ldots, x_m \in K$,*

$$N(\mathrm{conv}\{x_i\}_{i=1}^m, \frac{1}{2}T) \leq m^{(16T_p)^{p^*}}$$

(where $p^* = p/(1-p)$).

Proof Assume first that K has type p, and denote $K' = \mathrm{conv}\{x_i\}_{i=1}^m$. Then Lemma 1.5.3 implies that for any k

$$N(K', 2k^{-1/p^*}T_p(X)K) \leq m^k.$$

Choosing $k = (16T_p)^{p^*}$, so that $2k^{-1/p^*}T_p(X) = 1/8$, and using that $K \subset 4T$, we get that

$$N(K', \frac{1}{2}T) \leq m^{(16T_p)^{p^*}}.$$

For the other case, assume that T has type p and denote again $K' = \text{conv}\{x_i\}_{i=1}^m$. This time we think of K' as a subset of $4T$, thus $R = 4$ in Lemma 1.5.3, and we get that for any k

$$N(K', 8k^{-1/p^*}T_p(X)T) \leq m^k.$$

Choose again $k = (16T_p)^{p^*}$ so that $8k^{-1/p^*}T_p(X) = 1/2$. The proof of the corollary is complete. □

Proof of Lemma 1.5.2 The first inequality is obvious. As for the second, Maurey's Lemma in fact gives a much stronger inequality. Define yet another number

$$N'(K,T) = \min\{N : \exists x_1, \ldots, x_N \in K \ \text{ s.t. } \ K \subset \text{conv}\{x_i\} + T\}$$

Notice that $N'(K,T) \leq \hat{M}(K,T)$, because if one constructs inductively a maximal convexly separated set, it will satisfy the above condition and thus have cardinality greater than N'. Maurey's Lemma tells us that in the case $K \subset 4T$, for any $x_1, \ldots, x_m \in K$,

$$N(\text{conv}\{x_i\}_{i=1}^m, \frac{1}{2}T) \leq m^k,$$

where $k = (16T_p)^{p^*}$, and hence

$$N(\text{conv}\{x_i\}_{i=1}^m + \frac{1}{2}T, T) \leq m^k,$$

and so for $m < N(K,T)^{1/k}$ it cannot be that $K \subset \text{conv}\{x_i\}_{i=1}^m + \frac{1}{2}T$. Thus we have shown $N(K,T) \leq N'(K, \frac{1}{2}T)^k \leq \hat{M}(K, \frac{1}{2}T)^k$. □

1.5.4 Proof of duality for type $p > 1$ spaces

In this section we complete, as promised, the proof of the following theorem

Theorem 1.5.5 *There exists a universal constant c, and a constant $C_p(p, T_p)$ such that for any $p > 1$, and any two symmetric convex bodies K and T, such that either K or T has type p with constant T_p, one has*

$$N(K,T) \leq N(T^\circ, cK^\circ)^{C_p}.$$

Proof We apply Proposition 1.4.1, Proposition 1.5.1 and Lemma 1.5.2. Assume first that T is the body with type p and with type p constant T_p, and denote as usual $p^* = p/(1-p)$. Let L be any body with $L \subset 4T$. Then by Lemma 1.5.2 and Proposition 1.5.1

$$N(L,T) \leq \hat{M}(L, \frac{1}{2}T)^{(8T_p)^{p^*}} \leq \hat{M}(T^\circ, \frac{1}{4}L^\circ)^{2(8T_p)^{p^*}}.$$

Using the trivial side of Lemma 1.5.2 we get that

$$N(L,T) \leq N(T^\circ, \frac{1}{4}L^\circ)^{(8T_p)^{p^*}}.$$

Since this is true for *every* L with $L \subset 4T$, we use the first part of Proposition 1.4.1 to get that for every K

$$N(K,T) \leq \hat{M}(T^\circ, c'L^\circ)^{C_p}$$

(where $c' = 1/32$ and $C_p = 8(8T_p)^{p^*}$). The same argument holds if we assume, to begin with, that K is the one with type p and with type p constant T_p, and use the "Dually" part of Proposition 1.4.1. $\square$.

1.5.5 Some more on convex separation

In this section we will show three other results concerning covering numbers and convex separation.

The first result which we prove (in subsections 1.5.5.1 and 1.5.5.2), is that for two ellipsoids $\mathcal{E}_1$, $\mathcal{E}_2$, the usual covering number $N(\mathcal{E}_1, \mathcal{E}_2)$, and the maximal cardinality of a convexly separated set $\hat{M}(\mathcal{E}_1, \mathcal{E}_2)$, are equivalent up to a universal constant. As usual, we have a two step procedure, expanding the covering number with aid of equations (1.4.17), (1.4.18) (this time for ellipsoids) and then collapsing the large product with some telescoping argument. However, we need a new telescoping argument that will work for *convex* separation. This tends to be significantly more difficult than telescoping for usual covering. However, for *ellipsoids* it works.

The second result which we prove (in subsection 1.5.5.3) is the König-Milman type result, similar to Lemma 1.4.6, which follows from a more general fact about usual separation and convex separation: if a maximal separated set has cardinality which is exponentially large in the dimension, the same is true for a maximal convexly separated set, up to universal constants.

Thirdly, we present a result much in the spirit of [Tom], which is some 'integral' form of equivalence between M and $\hat{M}$ in the case where one of the bodies has some nontrivial type $p > 1$.

1.5.5.1 Telescoping $\hat{M}$

We begin by describing the variant of our telescoping argument, which is valid for convex separation, and how it can be applied to ellipsoids. We will assume here that $\hat{M}$ is defined via convexly separated sequences whose first element is 0. This is equivalent to the unrestricted version: If $x_1, \ldots, x_N$ is an ε-convexly separated set in K, then $(x_j - x_1)$ is an ε-convexly separated set in $2K$, so it is only a question of a factor 2 inside $\hat{M}$. Next, the inductive construction using convex combinations and lexicographic order preserves the property "the sequence starts from 0."

Lemma 1.5.6 *Let $K_1 \subset \mathbb{R}^{n_1}$, $K_2 \subset \mathbb{R}^{n_2}$ and let $K := \mathrm{conv}(K_1 \cup K_2) \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \equiv \mathbb{R}^n$ (where $n = n_1 + n_2$). Further, let $a > B > b > 0$ be such that $K_2 \subset BD$. Then*

$$\hat{M}(K_1, aD)\, \hat{M}(K_2, bD) \leq \hat{M}(K, \frac{b}{4}D).$$

Remark. Clearly, the parameter B is superfluous in the statement above. On the other hand, if we remove the hypothesis $a > B$ but keep $K_2 \subset BD$, the inequality holds if we replace $b/4$ in the right hand side by a quantity that is of order $\frac{a}{B}b$.

Proof of Lemma 1.5.6 Denote $M_1 := \hat{M}(K_1, aD)$, $M_2 := \hat{M}(K_2, bD)$. Let $0 = x_1, \ldots, x_{M_1} \in K_1$ and $0 = y_1, \ldots, y_{M_2} \in K_2$ be respectively a- and b-convexly separated. Further, for $i = 2, \ldots, M_1$ and $j = 2, \ldots, M_2$, let $\phi_i, \psi_j \in D$ be the corresponding separating functionals, as in the proof of Proposition 1.5.1. We thus have, for $i = 2, \ldots, M_1$ and $j = 2, \ldots, M_2$,

$$\phi_i(x_i) - a \geq \max_{l<i} \phi_i(x_l) \geq 0, \quad \psi_j(y_j) - b \geq \max_{k<j} \psi_j(x_k) \geq 0.$$

We now set $z_{ij} := \delta x_i + (1-\delta) y_j \in K$ and (for $i, j \geq 2$) $\xi_{ij} := \frac{(1-\delta)B\phi_i + \delta a \psi_j}{(1-\delta)B + \delta a} \in D$; δ will be chosen as $\frac{B}{a+B}$, but *a priori* it may be any number in $(0,1)$. If i or j (but not both) is 1, we set $\xi_{i1} := \phi_i$ and $\xi_{1j} := \psi_j$; these will be "trivial" cases. A direct calculation shows that if $j \geq 2$, then

$$\xi_{ij}(z_{lk}) = \frac{(1-\delta)\delta}{(1-\delta)B + \delta a} \left(B\phi_i(x_l) + a\psi_j(y_k) \right).$$

(Here we use that $\phi_i(y_k) = \psi_j(x_l) = 0$.) With our choice of δ the fraction on the right becomes $\alpha := (2(a+B))^{-1}$. We will verify that if $(i,j) > (l,k)$ (in the lexicographic order), then the difference between the second factors corresponding to $\xi_{ij}(z_{ij})$ and $\xi_{ij}(z_{lk})$ is at least ab; this will show that the z_{ij}'s are $\frac{ab}{2(a+B)} > \frac{b}{4}$ separated (remember that $B < a$), proving the lemma. To this end, note that if $k < j$ then

$$\alpha^{-1}\xi_{ij}(z_{ik}) = B\phi_i(x_i) + a\psi_j(y_k) \leq B\phi_i(x_i) + a(\psi_j(y_j) - b) = \alpha^{-1}\xi_{ij}(z_{ij}) - ab.$$

On the other hand, if $l < i$ (and $j \geq 2$), then

$$\begin{aligned}
\alpha^{-1}\xi_{ij}(z_{lk}) &= B\phi_i(x_l) + a\psi_j(y_k) \\
&= \alpha^{-1}\xi_{ij}(z_{ij}) - a(B - \psi_j(y_k)) - a(\psi_j(y_j) - b) - ab \\
&\leq \alpha^{-1}\xi_{ij}(z_{ij}) - ab,
\end{aligned}$$

again giving the required separation. Finally, if i or j equals 1 then the calculation is even simpler. E.g., in the latter case (with necessarily $l < i$),

$$\xi_{i1}(z_{lk}) = \phi_i(z_{lk}) = \delta\phi_i(x_l) \leq \delta(\phi_i(x_i) - a) = \xi_{i1}(z_{i1}) - \delta a,$$

and it remains to verify that under our assumptions $\delta a = \frac{Ba}{a+B} > \frac{b}{4}$. □

1.5.5.2 Completing the proof of equivalence for ellipsoids

We are now able to prove the following theorem.

Theorem 1.5.7 *There exist universal constants c'' and C'' such that for any two ellipsoids $\mathcal{E}_1, \mathcal{E}_2$ one has*

$$N(\mathcal{E}_1, \mathcal{E}_2) \leq \hat{M}(\mathcal{E}_1, c''\mathcal{E}_2)^{C''}.$$

We need the following simple lemma,

Lemma 1.5.8 *For any R one has*

$$\hat{M}(RD_k, D_k) \geq R^{k/2}.$$

Proof Take a maximal $\sqrt{2R}$-separated set on the sphere of radius R, it has cardinality $\geq R^{k/2}$, and each point is at distance at least 1 from the convex hull of all the others, hence the set is 1-convexly separated. □

Proof of Theorem 1.5.7 From the invariance of (all types of) covering numbers to invertible linear transformations, we may without loss of generality assume that $\mathcal{E}_2 = D$, and we denote $\mathcal{E}_1$ by $\mathcal{E}$.

We begin by estimating $\hat{M}(\mathcal{E}, D)$ from below (thus, the constant c'' will actually appear, in inverse, on the left hand side). We actually estimate $\hat{M}(5\mathcal{E}, 5D)$, and start with replacing $5\mathcal{E}$ by $\mathcal{E}'$ in the following way: for each axis λ_i of $5\mathcal{E}$, we give $\mathcal{E}'$ an axis $\mu_i = 5^{5^k} \leq \lambda_i$ for the largest k possible (we are assuming that all axes are greater than 5 to being with, discarding the rest only reduces this number, and if all axes are smaller than 5 we trivially have the theorem). We denote the number of i's for which $\mu_i = 5^{5^k}$ by Δ_k. Thus if we take the balls $5^{5^k} D_{\Delta_k}$ in $\mathbb{R}^{\Delta_k}$ and consider $\mathbb{R}^n = \prod_{k=0}^m \mathbb{R}^{\Delta_k}$ then

$$\mathcal{E} \supset \mathcal{E}' \supset \times_{k=0}^m 5^{5^k} D_{\Delta_k}.$$

Therefore, $\hat{M}(\mathcal{E}, tD) \geq \hat{M}(\times_{k=0}^m 5^{5^k} D_{\Delta_k}, tD)$. To estimate the latter we use Lemma 1.5.6. We use it with $a = B$ actually, but this can be avoided, (we omit the dimensions of the euclidean balls when unneeded)

$$\hat{M}(\times_{k=j}^m 5^{5^k} D_{\Delta_k}, 5^{5^{j-1}} D) \quad \geq \quad \hat{M}(\times_{k=j+1}^m 5^{5^k} D_{\Delta_k}, 5^{5^j} D) \hat{M}(5^{5^j} D_{\Delta_j}, 4 \times 5^{5^{j-1}} D).$$

Thus we get that

$$\hat{M}(\times_{k=1}^m 5^{5^k} D_{\Delta_k}, 5D) \quad \geq \quad \prod_{j=1}^m \hat{M}(5^{5^j} D_{\Delta_j}, 4 \times 5^{5^{j-1}} D).$$

Now, however, we are in the euclidean case, where we can use Lemma 1.5.8. The above implies

$$\hat{M}(\times_{k=1}^m 5^{5^k} D_{\Delta_k}, 5D) \quad \geq \quad \prod_{j=1}^m \hat{M}(5^{5^{j-1} \cdot 2} D_{\Delta_j}, D) \geq \prod_{j=1}^m 5^{5^{j-1} \Delta_j}.$$

We turn now to the left hand side. On the left hand side we simply have $N(\mathcal{E}, tD)$ (t to be later decided), and for this number there exist well known estimates. In particular, since we need an estimate from above, we use the following: assume that the axes of $\mathcal{E}$ are $\lambda_1 \geq \ldots \geq \lambda_n$. We disregard all axes such that $\lambda_i < t/2$, so we have a lower dimensional ellipsoid $\mathcal{E}''$ for which, in particular $(t/2)D \subset \mathcal{E}''$ and such that $N(\mathcal{E}'', (t/2)D) \geq N(\mathcal{E}'' + (t/2)D, tD) \geq N(\mathcal{E}, tD)$. So we see that it is enough to show that for $\mathcal{E}$ with $\frac{t}{2} D \subset \mathcal{E}$, we have $N(\mathcal{E}, (t/2)D)$ small. We remark that $\hat{M}(\mathcal{E}, D)$ is only smaller if you restrict to a subspace, and that in the special case where all axes are smaller than $t/2$ we trivially have the inequality since the left hand number is 1. We now estimate

$$N(\mathcal{E}'', (t/2)D) \leq \frac{\mathrm{Vol}(\mathcal{E} + \frac{t}{4} D)}{\mathrm{Vol}(\frac{t}{4} D)} \leq \frac{\mathrm{Vol}((3/2)\mathcal{E})}{\mathrm{Vol}(\frac{t}{4} D)} \leq \prod_{i=1}^m ((\frac{6}{t}) \times 5^{5^k})^{\Delta_k}.$$

We need thus to compare $\prod_{k=1}^m (\frac{6}{t} \times 5^{5^k})^{\Delta_k}$ with $\prod_{j=1}^m 5^{5^{j-1} \Delta_j}$, which are clearly comparable, once we choose $t = 30$ say, up to a power 5. Note that our discarding axes smaller than 5 can be compensated by the choice of a large enough t.
□

1.5.5.3 A König-Milman type Result

We want to recover a version of the result of König and Milman [KöM], which was formulated as Lemma 1.4.6 above. Their result gives duality for two general convex bodies in the case where the covering number is exponential in the dimension. The standard way to prove this is using volume estimates. We will give a different proof, using convex separation. It is a generalization of their result in the following sense: equation (1.4.13) gives duality in case $N(K,T) \geq 2^{cn}$ for a *large* c, and Corollary 1.5.13 below will give duality in case $N(K,T) \geq 2^{cn}$ for *any* c (with constants of duality depending, unfortunately, on c; but uniformly bounded from below if c is). Notice, however, that we are forced to include an additional constant next to the body. This cannot be avoided, if one considers the example of $N(B_\infty^n, (1-1/n)B_\infty^n)$ and $N(B_1^n, (1-1/n)B_1^n)$, the first is 2^n, exponentially large, and the second is not as large.

The first step, which is a result on its own right, will be to show that if the covering number is exponential in the dimension of the space, then so is the cardinality of a maximal convexly separated set. Then one can simply use the duality for this number (proved in Section 1.5.2 above) to obtain the result of König-Milman type. Thus our goal is thus to prove the following Theorem

Theorem 1.5.9 *Let K and T be two centrally symmetric convex bodies in R^n, and assume that $N(K,T) \geq 2^{cn}$ for some constant c. Then*

$$\hat{M}(K, c''T) \geq 2^{c''n}.$$

where $c'' = c''(c)$ depends only on c.

We use a special case of a result of Pisier [Pis2] concerning M-ellipsoids. The result reads as follows

Lemma 1.5.10 *For every $\alpha < 2$ there exists a constant $c(\alpha)$ such that for every convex body $K \subset R^n$ there exists an ellipsoid $\mathcal{E}_K$ with the following property: for every $t > 0$,*

$$\max\{N(K, t\mathcal{E}_K), N(\mathcal{E}_K, tK), N(K^\circ, t(\mathcal{E}_K)^\circ), N((\mathcal{E}_K)^\circ, tK^\circ)\} \leq 2^{c(\alpha)n/t^\alpha}.$$

We will use this result for example with $\alpha = 1$, and for convenience denote the constant by c_1. We will also need the following easy inequality, similar to Fact 1.3.3

Fact 1.5.11 *For any convex symmetric A, B and C,*

$$\hat{M}(A, C) \leq N(A, B)\hat{M}(B, C). \tag{1.5.6}$$

Proof Indeed, if $\{y_1, \dots, y_M\}$ are convexly separated in A with respect to C, and if A is covered by N copies of B, then in each translate $x_i + B$ the set $\{y_1, \dots, y_M\} \cap (x_i + B)$ is convexly separated, and at least one of them has no less than M/N elements. □

With the aid of Proposition 1.5.1, we can also show the following fact, which we need as a technical tool later on in this section

Fact 1.5.12 *For any convex symmetric A, B and C,*

$$\hat{M}(A, C) \le \hat{M}(A, B)^4 N(C^\circ, \frac{1}{4}B^\circ)^2. \tag{1.5.7}$$

Proof We write the following line of inequalities, the first coming from duality of convex separation, and the second being Fact 1.5.11:

$$\begin{aligned} \hat{M}(A, C) &\le \hat{M}(C^\circ, \frac{1}{2}A^\circ)^2 \le (N(C^\circ, \frac{1}{4}B^\circ)\hat{M}(B^\circ, 2A^\circ))^2 \\ &\le N(C^\circ, \frac{1}{4}B^\circ)^2(\hat{M}(A, B) + 1)^4. \end{aligned}$$

□

Proof of Theorem 1.5.9 We are assuming that $N(K, T) \ge 2^{cn}$. Fact 1.3.3 gives that

$$N(K, T) \le N(K, t\mathcal{E}_K)N(t\mathcal{E}_K, \frac{1}{t}\mathcal{E}_T)N(\mathcal{E}_T, tT)$$

for any $t > 0$. If we choose $t = 2c_1/c$, and use Pisier's Lemma 1.5.10, we get that $N(\mathcal{E}_K, \frac{1}{t^2}\mathcal{E}_T) \ge 2^{cn/2}$. However, for ellipsoids we have already showed in Section 1.5.5.2 the equivalence of covering and convex packing up to universal constants. Thus

$$\hat{M}(\mathcal{E}_K, c'\mathcal{E}_T) \ge 2^{c'n}$$

(we use the constant c' to denote some constant depending only on c). We now use Fact 1.5.11, which gives that

$$\hat{M}(\mathcal{E}_K, c'\mathcal{E}_T) \le N(\mathcal{E}_K, sK)\hat{M}(sK, c'\mathcal{E}_T),$$

and continue with Fact 1.5.12 to get (for any s) that

$$\hat{M}(\mathcal{E}_K, c\mathcal{E}_T) \le N(\mathcal{E}_K, sK)\hat{M}(sK, \frac{c'}{4s}T)^4 N(T, s\mathcal{E}_T)^2$$

We choose s so that the two terms do not get in the way, that is, $s = 4\frac{c_1}{c'}$, and finally end up, for c'' depending only on c, with $\hat{M}(K, c''T) \ge 2^{c''n}$. □

We can now use the duality established for $\hat{M}$, that is, Proposition 1.5.1, to prove the following König-Milman-type result. It is worse than (1.4.13) since there is a constant next to the body, but better since it applies for any $c > 0$, whereas König-Milman's result (1.4.13) will give the below only for $c > c_0$ for some c_0. This is explained in detail in Chapter 3, Section 3.2.6, where we derive a similar estimate *in a special case* by simpler methods.

Corollary 1.5.13 *There exists a function $c'(c)$ such that the following holds. For any two symmetric convex bodies K and T in $\mathbb{R}^n$, such that $N(K,T) \geq 2^{cn}$, on has*

$$N(T^\circ, c'K^\circ) \geq 2^{c'n}.$$

Proof We use Theorem 1.5.9 to say that if $N(K,T) \geq 2^{cn}$ then

$$\hat{M}(K, c''T) \geq 2^{c''n}.$$

with $c'' = c''(c)$. The duality of convex separation, Proposition 1.5.1, implies that

$$\hat{M}(T^\circ, \frac{c''}{2}K^\circ) \geq 2^{\frac{c'}{2}n},$$

and this of course means that also

$$N(T^\circ, \frac{c''}{2}K^\circ) \geq 2^{\frac{c'}{2}n}.$$

□

1.5.5.4 A global form of equivalence

Although for the time being in order to show that $M(K,T)$ and $\hat{M}(K,T)$ are equivalent we need (for example) that $K \subset RT$ *and* that one of the two bodies has non trivial type (and the equivalence constants depend on R and p, T_p), we do have a way to get some weaker form of equivalence *without* the boundedness condition. This is very much in the spirit of the papers [Tom] and [BPST]. The main step is the following lemma (which is a direct consequence of the arguments in the above sections).

Lemma 1.5.14 *Let K and T be two convex bodies, one of which has some non-trivial type p with type constant T_p. Then for any $\varepsilon > 0$*

$$N(K, \varepsilon T) \leq N(K, 2\varepsilon T)\hat{M}(K, \frac{\varepsilon}{4}T)^{(16T_p)^{p/(p-1)}}.$$

Proof First assume that the body T has type p. Consider the following inequality which is a consequence of Facts 1.3.3 and 1.3.5.

$$N(K, \varepsilon T) \leq N(K, 4\varepsilon T \cap 2K)N(4\varepsilon T \cap 2K, \varepsilon T) \leq N(K, 2\varepsilon T)N(4\varepsilon T \cap 2K, \varepsilon T).$$

The last number is a covering number, where the first body is contained in 4 times the second, and the second has type p with type constant T_p. Therefore Lemma 1.5.2 tells us that $N(4\varepsilon T \cap 2K, \varepsilon T) \le \hat{M}(4\varepsilon T \cap 2K, (\varepsilon/2)T)^k$ for $k = (16T_p)^{p/(p-1)}$. Thus we have shown

$$N(K, \varepsilon T) \le N(K, 2\varepsilon T)\hat{M}(4\varepsilon T \cap 2K, (\varepsilon/2)T)^{(16T_p)^{p/(p-1)}},$$

from which the conclusion of the lemma follows. In the case where K has type p we consider the line of inequalities

$$\begin{aligned} N(K, \varepsilon T) &\le N(K, \mathrm{conv}(\frac{\varepsilon}{2}T \cup \frac{1}{4}K))N(\mathrm{conv}(\frac{\varepsilon}{2}T \cup \frac{1}{4}K), \varepsilon T) \\ &\le N(K, \mathrm{conv}(\frac{\varepsilon}{2}T \cup \frac{1}{4}K))N(\frac{\varepsilon}{2}T + \frac{1}{4}K, \varepsilon T) \\ &\le N(K, \mathrm{conv}(\frac{\varepsilon}{2}T \cup \frac{1}{4}K))N(K, 2\varepsilon T), \end{aligned}$$

which follow from Facts 1.3.3 and 1.3.4, and from the inclusion $\mathrm{conv}(A, B) \subset A+B$. We may now use Lemma 1.5.2 again, to see that $N(K, \mathrm{conv}(\frac{\varepsilon}{2}T \cup \frac{1}{4}K)) \le \hat{M}(K, \mathrm{conv}(\frac{\varepsilon}{4}T \cup \frac{1}{8}K))^k$ for $k = (16T_p)^{p/(p-1)}$, which implies

$$N(K, \varepsilon T) \le N(K, 2\varepsilon T)\hat{M}(K, \mathrm{conv}(\frac{\varepsilon}{4}T \cup \frac{1}{8}K))^{(16T_p)^{p/(p-1)}}.$$

The conclusion of Lemma 1.5.14 follows. □

Applications of Lemma 1.5.14 can be of various natures. For example we may show the following

Proposition 1.5.15 *For every $\alpha > 0$, $1 < p \le 2$ and $T_p \ge 1$ there exists a constant $C = C(\alpha, p, T_p)$ such that if either K or T has type p with type constant T_p then*

$$\sup_\varepsilon \varepsilon^\alpha \log \hat{M}(K, \varepsilon T) \le \sup_\varepsilon \varepsilon^\alpha \log N(K, \varepsilon T) \le C \sup_\varepsilon \varepsilon^\alpha \log \hat{M}(K, \varepsilon T).$$

The constant can be taken to be $C(\alpha, p, T_p) = (16T_p)^{p/(p-1)}4^\alpha(1 - (1/2)^\alpha)^{-1}$.

Proof the first inequality is trivial, since $\hat{M} \le N$ always. Lemma 1.5.14 tells us that for any α

$$\varepsilon^\alpha \log N(K, \varepsilon T) \le \varepsilon^\alpha \log N(K, 2\varepsilon T) + \varepsilon^\alpha \log \hat{M}(K, \frac{\varepsilon}{4}T)^{(16T_p)^{p/(p-1)}}.$$

Therefore

$$\sup_\varepsilon \varepsilon^\alpha \log N(K, \varepsilon T) \le \sup_\varepsilon (\varepsilon/2)^\alpha \log N(K, \varepsilon T) + (16T_p)^{p/(p-1)} \sup_\varepsilon \varepsilon^\alpha \log \hat{M}(K, \frac{\varepsilon}{4}T),$$

and we conclude that

$$(1-(1/2)^{\alpha})\sup_{\varepsilon}\varepsilon^{\alpha}\log N(K,\varepsilon T)\leq(16T_p)^{p/(p-1)}4^{\alpha}\sup_{\varepsilon}\varepsilon^{\alpha}\log\hat{M}(K,\varepsilon T).$$

The proof of the proposition is complete. □

This conclusion is perhaps better expressed in terms of entropy numbers. Recall that the entropy numbers of a linear operator $u : X \to Y$ are defined by

$$e_k(u) := \inf\{\varepsilon : N(uB(X),\varepsilon B(Y)) \leq 2^{k-1}\},$$

where $B(X)$ and $B(Y)$ are the unit balls of X and Y respectively. We may define, similarly, convexified entropy numbers by

$$\hat{e}_k(u) := \inf\{\varepsilon : \hat{M}(uB(X),\varepsilon B(Y)) \leq 2^{k-1}\}.$$

We trivially have that $\hat{e}_k(u) \leq e_k(u)$. The proposition can be rephrased as (where $C'(\alpha,p,T_p)$ is $C(1/\alpha,p,T_p)^{\alpha}$ from the above proposition):

Proposition 1.5.16 *For any $\alpha > 0$, p and T_p there exists a constant $C'(\alpha,p,T_p)$ such that for any operator $u : X \to Y$, where either X or Y has type p with type constant T_p one has*

$$\sup_k k^{\alpha}\hat{e}_k(u) \leq C'(\alpha,p,T_p)\sup_k k^{\alpha}e_k(u).$$

Similarly we can show that the sequence $e_k(u)$ is in some ℓ_q space if and only $\hat{e}_k(u)$ is (with the condition that either the domain or the range of u has some non trivial type p). Namely we prove

Proposition 1.5.17 *Let $u : X \to Y$ be a linear operator, and assume that either X or Y has some non trivial type $p > 1$. Then for any $q > 0$ we have*

$$\sum_{k=1}^{\infty} e_k^q(u) < \infty \quad \text{if and only if} \quad \sum_{k=1}^{\infty}\hat{e}_k^q(u) < \infty.$$

Proof We need only show one direction. Consider first the following integral inequality, for any $q > -1$

$$\int_0^{\infty}\varepsilon^q\log N(K,\varepsilon T)d\varepsilon \leq C\int_0^{\infty}\varepsilon^q\log\hat{M}(K,\varepsilon T)d\varepsilon, \tag{1.5.8}$$

where C depends only on q, p and T_p. This inequality can be verified directly: Lemma 1.5.14 implies that

$$\begin{aligned}\int_0^\infty \varepsilon^q \log N(K,\varepsilon T)d\varepsilon &\leq \int_0^\infty \varepsilon^q \log N(K,2\varepsilon T)d\varepsilon \\ &+ C_{p,T_p}\int_0^\infty \varepsilon^q \log \hat{M}(K,\frac{1}{4}\varepsilon T)d\varepsilon \\ &= (1/2)^{q+1}\int_0^\infty \varepsilon^q \log N(K,\varepsilon T)d\varepsilon \\ &+ 4^{q+1}C_{p,T_p}\int_0^\infty \varepsilon^q \log \hat{M}(K,\varepsilon T)d\varepsilon.\end{aligned}$$

This means that

$$\int_0^\infty \varepsilon^q \log N(K,\varepsilon T)d\varepsilon \leq \left(4^{q+1}C_{p,T_p}(1-(1/2)^{q+1})^{-1}\right)\int_0^\infty \varepsilon^q \log \hat{M}(K,\varepsilon T)d\varepsilon.$$

The inequality (1.5.8) can be translated into an inequality for the ℓ_{q+1} norms of the sequences. Indeed, denote $K = B(X)$ and $T = uB(Y)$, we have

$$\int_0^\infty \varepsilon^q \log N(K,\varepsilon T)d\varepsilon = \sum_{k=0}^\infty \int_{e_{k+1}}^{e_k} \varepsilon^q \log N(K,\varepsilon T).$$

Hence

$$\begin{aligned}\int_0^\infty \varepsilon^q \log N(K,\varepsilon T)d\varepsilon &\leq \sum_{k=0}^\infty \int_{e_{k+1}}^{e_k} \varepsilon^q k &= \\ \sum_{k=0}^\infty \frac{1}{q+1}(e_k^{q+1} - e_{k+1}^{q+1})k &= \frac{1}{q+1}\sum_{k=1}^\infty e_k^{q+1}.\end{aligned}$$

Similarly

$$\begin{aligned}\int_0^\infty \varepsilon^q \log \hat{M}(K,\varepsilon T)d\varepsilon &\leq \sum_{k=1}^\infty \int_{\hat{e}_{k+1}}^{\hat{e}_k} \varepsilon^q (k-1) &= \\ \sum_{k=1}^\infty \frac{1}{q+1}(\hat{e}_k^{q+1} - \hat{e}_{k+1}^{q+1})(k-1) &= \frac{1}{q+1}\sum_{k=2}^\infty \hat{e}_k^{q+1}.\end{aligned}$$

Therefore knowing that the integrals are equivalent up to a universal constant means also that one of the sequences is in ℓ_{q+1} if and only if the other one is, and the assertion is proved. □

Finally, we mention that the ℓ_∞ norms of the sequences are trivially identical. Indeed, it is worthwhile to notice the simple fact that we always have $e_1 = \hat{e}_1$ and $e_2 = \hat{e}_2$ (and the sequences are decreasing).

Chapter 2

Shannon Entropy

2.1 Introduction, Definitions, Statement of Results

This chapter includes the results concerning information theoretic entropy, also called "Shannon Entropy". It contains results from the papers [ABBN1] and [ABBN2]. In this section we give the needed definitions, some introductory remarks, and state the main results.

The Shannon entropy of a real valued random variable X with density $f : \mathbb{R} \to [0, \infty)$ is defined as

$$\mathrm{Ent}(X) = \int_{\mathbb{R}} -f \log f$$

provided the positive part of the integral is finite (so $\mathrm{Ent}(X) \in [-\infty, +\infty)$). If X has variance 1, it is a classical fact that its entropy is well defined and bounded above by that of a standard Gaussian random variable G. Thus, among random variables with variance 1, the standard Gaussian G has the largest entropy. If X_i are independent copies of a random variable X with variance 1, then the normalized sums

$$S_n = \frac{1}{\sqrt{n}} \sum_1^n X_i$$

approach the standard Gaussian as n tends to infinity, in a variety of senses.

In the 1940's Shannon proved that $\mathrm{Ent}(S_2) \geq \mathrm{Ent}(S_1)$, that is, the entropy of the normalized sum of two independent copies of a random variable is larger than that of the original (for proofs of this fact see [ShWe, Sta, Lie]). Inductively this implies that for the sequence of powers of 2, entropy is increasing, i.e., that

$\mathrm{Ent}(S_{2^k}) \geq \mathrm{Ent}(S_{2^{k-1}})$. It was therefore naturally conjectured that the entire sequence $\mathrm{Ent}_n = \mathrm{Ent}(S_n)$ increases with n. This problem, though seemingly elementary, has resisted the attempts of many mathematicians, and was formally put forth as a conjecture by Lieb in 1978 [Lie]. It is also closely related to the work of Linnik, see [Lin]. For example, it was unknown even whether the relation $\mathrm{Ent}_2 \leq \mathrm{Ent}_3$ holds true. The difficulty arises from the fact that there is no natural way to "build" the sum of three independent copies of X out of copies of the sums of two.

One of the two main results to be presented in this chapter is a solution of this problem. We develop a new understanding of the underlying structure of sums of independent random variables and use it to prove the general statement for entropy:

Theorem 2.1.1 (Entropy increases at every step) *Let $X_1, X_2, \ldots$ be independent and identically distributed square-integrable random variables. Then for every n one has*

$$\mathrm{Ent}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) \leq \mathrm{Ent}\left(\frac{X_1 + \cdots + X_{n+1}}{\sqrt{n+1}}\right).$$

The proof of Theorem 2.1.1 is presented in Section 2.3. A by-product of our approach is the following generalization of the so called *Shannon entropy power inequality*, the case $n = 2$ of which is a classical fact (see [Lie]).

Theorem 2.1.2 (Entropy power for many summands) *Let $X_1, X_2, \ldots, X_n$ be independent square-integrable random variables. Then*

$$\exp\left(2\,\mathrm{Ent}\sum_{i=1}^{n} X_i\right) \geq \frac{1}{n-1}\sum_{j=1}^{n}\exp\left(2\,\mathrm{Ent}\sum_{i\neq j}^{n} X_i\right).$$

This Theorem is deduced from a weighted version of Theorem 2.1.1. We explain this in Section 2.3.1

Our second main result to be presented in this section concerns the rate at which entropy is produced by linear combinations of independent random variables which satisfy a spectral gap condition. Before stating our theorem, a few words of introduction will be needed. By the Pinsker-Csiszar-Kullback inequality ([Pin], [Csi], [Kul]), G is a stable maximizer of entropy in the following strong sense:

$$d_{TV}(X - \mathbb{E}X, G)^2 \leq 2[\mathrm{Ent}(G) - \mathrm{Ent}(X)], \tag{2.1.1}$$

where d_{TV} is the total variation distance. It is therefore natural to track the convergence in the central limit theorem (CLT) in terms of entropy. Let $(X_n)_{n\in\mathbb{N}}$

be a sequence of independent copies of a random variable X. For notational convenience we assume that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$. Set, as before,

$$S_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

Linnik [Lin] used entropy to reprove convergence in distribution of S_n to G. He actually showed that $\mathrm{Ent}(S_n)$ converges to $\mathrm{Ent}(G)$ under some smoothness assumptions on the law of X. This is discussed in [Bar] where Barron established entropic convergence in the CLT in full generality: if the entropy of S_n ever becomes different from $-\infty$, then $\mathrm{Ent}(S_n)$ converges to $\mathrm{Ent}(G)$. This convergence is very strong, as (2.1.1) shows.

The fact that the sequence of entropies $\mathrm{Ent}(S_n)$ has a limit followed already from the Shannon-Stam inequality ([ShWe], [Sta]): for independent variables Y, Z and $\lambda \in (0,1)$ one has

$$\mathrm{Ent}(\sqrt{1-\lambda}\,Y + \sqrt{\lambda}Z) \geq (1-\lambda)\mathrm{Ent}(Y) + \lambda\mathrm{Ent}(Z), \tag{2.1.2}$$

therefore $(n\mathrm{Ent}(S_n))$ is super-additive and converges to its supremum. (Actually, as stated in Theorem 2.1.1, we have showed in [ABBN1] that the entire sequence sequence $(\mathrm{Ent}(S_n))$ is non-decreasing.)

Until recently, quantitative improvements on the Shannon-Stam inequality (2.1.2), have been somewhat elusive. A *qualitative* result in this direction, under pretty general hypotheses was obtained by Carlen and Soffer [CaSo] (see also the works of Brown and Shimuzu [Bro], [Shi]). For a special class of random variables X, quantitative estimates were obtained in [BBN] for the entropy jump occurring when one sums *two* independent copies of the random variable. More precisely, we say that the random variable X satisfies a Poincaré (or spectral gap) inequality with constant $c > 0$ if for every smooth function s, one has

$$c\,\mathrm{Var}(s(X)) \leq \mathbb{E}[(s'(X))^2]. \tag{2.1.3}$$

The main result in [BBN] asserts that independent copies X_1, X_2 of such a random variable X satisfy

$$\mathrm{Ent}\left(\frac{X_1 + X_2}{\sqrt{2}}\right) - \mathrm{Ent}(X) \geq \frac{c}{2+2c}(\mathrm{Ent}(G) - \mathrm{Ent}(X)).$$

In Section 2.4 we apply an extension of the variational method of [BBN] to estimate the entropy of sums (or linear combinations) of more than two copies of a random variable (this does not follow formally from the above). Independently of [BBN] and almost simultaneously, Barron and Johnson proved an analogous but slightly weaker fact for the Fisher information. We have just learned that, by modifying their argument they are able to recover our results for entropy: the new argument appears in [BaJo].

Our principal result in Section 2.4 is the following.

Theorem 2.1.3 *Let $X_1, \ldots, X_n$ be independent copies of a random variable X with density f. Assume that* $\mathrm{Var}(X) = 1$ *and that X satisfies the Poincaré inequality with constant c. Then for every $a \in \mathbb{R}^n$ with $\sum_{i=1}^n a_i^2 = 1$,*

$$\mathrm{Ent}(G) - \mathrm{Ent}\left(\sum_{i=1}^n a_i X_i\right) \leq \frac{\alpha(a)}{c/2 + (1-c/2)\alpha(a)}\left(\mathrm{Ent}(G) - \mathrm{Ent}(X)\right),$$

where $\alpha(a) = \displaystyle\sum_{i=1}^n a_i^4$.

Remark In fact, we obtain a slightly stronger, more complicated inequality, see inequality (2.4.13).

Note that when one of the a_i's tends to ± 1, there can be no increase of entropy. This is consistent with the fact that $\alpha(a)$ tends to 1 in this case. On the other hand, for equal coefficients $1/\sqrt{n}$ we obtain that the entropy distance from the Gaussian decays at rate $1/n$

$$\mathrm{Ent}(G) - \mathrm{Ent}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right) \leq \frac{1}{1 + \frac{c}{2}(n-1)}\left(\mathrm{Ent}(G) - \mathrm{Ent}(X)\right). \qquad (2.1.4)$$

In conjunction with the Csiszar-Kulback-Pinsker inequality, the above theorem leads to a quantitative information-theoretic proof of the central limit theorem (under a spectral gap assumption). The rate of convergence obtained (in total variation distance) is $O(1/\sqrt{n})$, which is well known to be optimal. Moreover, we also obtain Berry-Esseén type estimates, but in total variation, for the rate of convergence to a Gaussian.

This Chapter is organized as follows. The next section contains two important tools to be used in the proof of both Theorems. The first is a well known reduction, relating questions about entropy to questions about Fisher information. The second is new, and is a variational representation of Fisher information, which turns our to be extremely useful. The two dimensional case was proved in [BBN] and the general case in [ABBN1]. In Section 2.3 we use it to prove Theorem 2.1.1 (Shannon's monotonicity Conjecture) and the entropy power inequality Theorem 2.1.2. Section 2.4 is devoted to the proof of Theorem 2.1.3, the tracking down of rates of convergence in the Central Limit Theorem.

2.2 Main Tools

2.2.1 First tool - Reduction to inequalities for Fisher information

All our arguments begin with a reduction from entropy to another information theoretic notion, the Fisher information of a random variable. For sufficiently regular densities, the Fisher information can be defined by

$$J(X) = \int_{\mathbb{R}} \frac{(f')^2}{f}.$$

Among random variables with variance 1, the Gaussian has *smallest* information: $J(G) = 1$. There is a remarkable connection between Fisher information and entropy, provided by the adjoint Ornstein-Uhlenbeck semigroup, which goes back to de Bruijn, Bakry-Emery and Barron. According to Bruijn's identity (see e.g. [Sta]), Fisher information is the time derivative of entropy along the heat semi-group. A similar relation holds along the adjoint Ornstein-Uhlenbeck semigroup (see [BaEm, Bar]), where the evolute at time t of a variable X, denoted $X^{(t)}$, has the same law as $e^{-t}X + \sqrt{1-e^{-2t}}\,G$ where G is a normal Gaussian variable independent of X. More precisely, if $\mathrm{Ent}(X)$ is finite, then for every $t > 0$, $X^{(t)}$ has finite Fisher information and

$$\mathrm{Ent}(G) - \mathrm{Ent}(X) = \int_0^\infty \Big(J(X^{(t)}) - 1 \Big)\, dt.$$

A very clear explanation of this appears in [CaSo]. Since the action of the semigroup commutes with self-convolution, the evolute $(\sum a_i X_i)^{(t)}$ has the same distribution as $\sum a_i X_i^{(t)}$ (and in particular $(S_n(X))^{(t)}$ and $S_n(X^{(t)})$ have the same distribution).

The above implies, for example, that instead of proving Theorem 2.1.1, we can instead prove the following

Theorem 2.2.1 *Let $X_1, X_2, \ldots$ be independent and identically distributed square-integrable random variables. Then for every n one has*

$$J\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) \geq J\left(\frac{X_1 + \cdots + X_{n+1}}{\sqrt{n+1}}\right).$$

We will then integrate along the semigroup, and retrieve Theorem 2.1.1. Notice that since the inequality we have to prove is linear, it simply integrates.

Similarly, instead of proving Theorem 2.1.3, it suffices to prove the following

Theorem 2.2.2 *Let Y be a random variable with* $\mathrm{Var}(Y) = 1$, *finite entropy, and spectral gap at least $c > 0$. Let $t > 0$, $X = Y^{(t)}$ and $X_1, \ldots, X_n$ be independent copies of X. Then for every $a \in \mathbb{R}^n$ with $\sum_{i=1}^n a_i^2 = 1$,*

$$J\left(\sum_{i=1}^n a_i X_i\right) - 1 \leq \frac{\alpha(a)}{c/2 + (1 - c/2)\alpha(a)} \left(J(X) - 1\right).$$

Again, we will integrate this linear inequality along the semigroup to get Theorem 2.1.3.

Let us also point out that the hypotheses in the latter theorem are far from optimal; we could certainly get away with less smoothness for X than is guaranteed by its being an Ornstein-Uhlenbeck evolute. However the hypothesis is sufficient for our goal and ensures that the density f of X has an integrable second derivative (see e.g. the Appendix of [BBN]).

2.2.2 Second tool - Variational representation of information

To prove Theorems 2.2.1 and 2.2.2 above, which both concern Fisher information, we use a new and flexible formula for the Fisher information of marginals of distributions. This is described in Theorem 2.2.3. A version of this theorem (in the case $n = 2$) was introduced in [BBN], and was motivated by an analysis of the transportation of measures and a local analogue of the Brunn-Minkowski theorem. In our setting we will usually consider the distribution $w_n : \mathbb{R}^n \to (0, \infty)$ defined by

$$w_n(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n)$$

(we will sometimes denote it simply by w). Notice that if X has density f then the density of S_n, the normalized sum of n i.i.d. copies X_i of X, is the marginal of w_n in direction $\frac{1}{\sqrt{n}}(1, \ldots, 1)$. The density of the random variable $\sum a_i X_i$, for a unit vector a, is the marginal of w_n in direction a. The following theorem applies in a more general situation.

Theorem 2.2.3 (Characterization of the information of marginals) *Let $w : \mathbb{R}^n \to (0, \infty)$ be a continuously twice differentiable density on $\mathbb{R}^n$ with*

$$\int \frac{\|\nabla w\|^2}{w}\,, \int \|Hess(w)\| \; < \infty.$$

Let $e \in \mathbb{R}^n$ be a unit vector and h the marginal density in direction e defined by

$$h(t) = \int_{te + e^{\perp}} w.$$

Then the Fisher information of the random variable with density h (denoted $J(h)$) satisfies

$$J(h) \leq \int \left(\frac{\operatorname{div}(pw)}{w}\right)^2 w, \tag{2.2.1}$$

for any continuously differentiable vector field $p : \mathbb{R}^n \to \mathbb{R}^n$ with the property that for every x,

$$\langle p(x), e \rangle = 1$$

(and, say, $\int \|p\| w < \infty$). If w satisfies $\int \|x\|^2 w(x) < \infty$, then there is equality for some suitable vector field p.

Remark The condition $\int \|p\| w < \infty$ is not important in applications but makes for a cleaner statement of the theorem. We are indebted to Giovanni Alberti for pointing out this simplified formulation of the result.

Proof The conditions on w ensure that for each t,

$$h'(t) = \int_{te+e^\perp} \partial_e w. \tag{2.2.2}$$

If

$$\int \left(\frac{\operatorname{div}(pw)}{w}\right)^2 w$$

is finite, then $\operatorname{div}(pw)$ is integrable on $\mathbb{R}^n$ and hence on almost every hyperplane perpendicular to e. If $\int \|p\| w$ is also finite, then on almost all of these hyperplanes the $(n-1)$-dimensional divergence of pw in the hyperplane, integrates to zero by the Gauss-Green Theorem. Since the component of p in the e direction is always 1, we have that for almost every t,

$$h'(t) = \int_{te+e^\perp} \operatorname{div}(pw).$$

Therefore

$$J(h) = \int \frac{h'(t)^2}{h(t)} = \int \frac{\left(\int \operatorname{div}(pw)\right)^2}{\int w}$$

and by the Cauchy-Schwarz inequality this is at most

$$\int_{\mathbb{R}^n} \frac{(\operatorname{div}(pw))^2}{w},$$

which completes the proof of the first part of the theorem.

To find a field p for which there is equality, we want to arrange that

$$\operatorname{div}(pw) = \frac{h'(t)}{h(t)} w,$$

since then we have

$$\int_{\mathbb{R}^n} \frac{(\operatorname{div}(pw))^2}{w} = \int_{\mathbb{R}^n} \frac{h'(t)^2}{h(t)^2} w = J(h).$$

Since we also need to ensure that $\langle p, e\rangle = 1$ identically, we construct p, separately on each hyperplane perpendicular to e, to be a solution of the equation

$$\operatorname{div}_{e^\perp}(pw) = \frac{h'(t)}{h(t)} w - \partial_e w. \tag{2.2.3}$$

Regularity of p on the whole space can be ensured by using a "consistent" method for solving the equation on the separate hyperplanes. Note that for each t, the conditions on w ensure that $\operatorname{div}_{e^\perp}(pw)$ is integrable on $te + e^\perp$, while the "constant" $\frac{h'(t)}{h(t)}$ ensures that the integral is zero.

The hypothesis $\int \|x\|^2 w(x) < \infty$ is needed only if we wish to construct a p for which $\int \|p\| w < \infty$. For each real t and each $y \in e^\perp$ set

$$F(t, y) = \frac{h'(t)}{h(t)} w(te + y) - \partial_e w(te + y).$$

The condition

$$\int \frac{\|\nabla w\|^2}{w} < \infty$$

shows that

$$\int \frac{F^2}{w} < \infty$$

and under the assumption $\int \|x\|^2 w(x) < \infty$ this guarantees that

$$\int |F| \cdot \|x\| < \infty. \tag{2.2.4}$$

As a result we can find solutions to our equation

$$\operatorname{div}_{e^\perp}(pw) = F$$

satisfying $\int \|pw\| < \infty$, for example in the following way.

Fix an orthonormal basis of $e^\perp$ and index the coordinates of $y \in e^\perp$ with respect to this basis, $y_1, y_2, \ldots, y_{n-1}$. Let $K(y_2, \ldots, y_{n-1}) = \int F(t, y)\, dy_1$ and choose a rapidly decreasing density g on the line. Then the function

$$F(t, y) - g(y_1) K(y_2, \ldots, y_n)$$

integrates to zero on every line in the y_1 direction so we can find $p_1(y)$ so that $p_1 w$ tends to zero at infinity and its y_1-derivative is $F(t, y) - g(y_1) K(y_2, \ldots, y_n)$.

The integrability of $p_1 w$ is immediate from equation (2.2.4). Continue to construct the coordinates of p, one by one, by conditioning on subspaces of successively smaller dimension. □

Remarks Theorem 2.2.3 could be stated in a variety of ways. The proof makes it clear that all we ask of the expression $\operatorname{div}(pw)$ is that it be a function for which $\operatorname{div}_{e^\perp}(pw) - \partial_e w$ integrates to zero on each hyperplane perpendicular to e. If pw decays sufficiently rapidly at infinity, then this condition is guaranteed by the Gauss-Green Theorem, and thus there is a certain naturalness about realizing the function as a divergence. However, the real point of the formulation given above is that in each of the applications that we have found for the theorem, it is much easier to focus on the vector field p than on its divergence, even though formally, the fields themselves play no role. This latter fact is reassuring given that there are so many different fields with the same divergence.

Formula (2.2.1) is very short, but it sometimes leads to tedious calculations when applied to specific p. For this reason we provide another formulation of it, which is sometimes easier to use.

Lemma 2.2.4 *Let w be as in Theorem 2.2.3 and let $p : \mathbb{R}^n \to \mathbb{R}^n$ be twice continuously differentiable, such that each coordinate function p_i is bounded and has bounded derivatives of first and second order. Then*

$$\int \frac{(\operatorname{div}(pw))^2}{w} = \int w\left[\operatorname{Tr}((Dp)^2) + \langle \operatorname{Hess}(-\log w)p, p\rangle\right]. \tag{2.2.5}$$

Proof of Lemma 2.2.4 Distributing the divergence and the square gives

$$\int \frac{(\operatorname{div}(pw))^2}{w} = \int \left(w(\operatorname{div}(p))^2 + \frac{\langle \nabla w, p\rangle^2}{w} + 2\operatorname{div}(p)\langle \nabla w, p\rangle \right).$$

On the other hand

$$\begin{gathered} \int w\left[\operatorname{Tr}((Dp)^2) + \langle \operatorname{Hess}(-\log w)p, p\rangle\right] = \\ \int \left(w\operatorname{Tr}((Dp)^2) + \frac{\langle \nabla w, p\rangle^2}{w} - \langle \operatorname{Hess}(w)p, p\rangle \right). \end{gathered}$$

So the identity we want amounts to

$$\int \left(w(\operatorname{div}(p))^2 + 2\operatorname{div}(p)\langle \nabla w, p\rangle - w\operatorname{Tr}((Dp)^2) + \langle \operatorname{Hess}(w)p, p\rangle \right) = 0.$$

This can be checked by integration by parts, once for the second term and twice for the fourth term, in order to get rid of derivatives of w. □

2.3 Proof of Shannon's Conjecture

In this section we complete the proof of Theorem 2.1.1. As explained in Section 2.2.1, in is enough to prove Theorem 2.2.1. Denote for convenience $J_n = J(S_n)$, the information of the normalized sum of n i.i.d. copies of X. As explained in the previous section, we shall consider densities of the type

$$w_n(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n),$$

where f is the one dimensional density of X. The density of S_n is thus the marginal of w_n in direction $e = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$.

From Theorem 2.2.3 above, we have an n-dimensional vector field $(p_1, \ldots, p_n) : \mathbb{R}^n \to \mathbb{R}^n$ with $< p, e >= 1$, that is, $\sum p_i = \sqrt{n}$, for which

$$J_n = \int_{\mathbb{R}^n} \left(\frac{\mathrm{div}(pw_n)}{w_n} \right)^2 w_n.$$

We shall use this optimizing n-dimensional vector field to construct an $(n+1)$-dimensional vector field $(P_1, \ldots, P_{n+1})$ which yields an upper bound for J_{n+1}. We then analyze this upper bound using a simple Hilbert space lemma on orthogonal projections, and show that it is at most J_n. The fact that the increase in entropy (or decrease of information) from $n = 1$ to $n = 2$ is so much easier to prove, is reflected in the fact that for $n = 1$, an optimizing p is constant.

Proof of Theorem 2.1.1 Choose an n-dimensional vector field $(p_1, \ldots, p_n) : \mathbb{R}^n \to \mathbb{R}^n$ with $\sum p_i = \sqrt{n}$ for which

$$J_n = \int_{\mathbb{R}^n} \left(\frac{\mathrm{div}(pw_n)}{w_n} \right)^2 w_n.$$

To construct a suitable vector field P on $\mathbb{R}^{n+1}$, we place a copy of the n-dimensional p onto each of the $(n+1)$ hyperplanes $\{x_i = 0\}$. One way to do this, consistent with the choice of p, is as follows. Set

$$\begin{aligned}
q_1 &= (0, p_1(x_2, \ldots, x_{n+1}), p_2(x_2, \ldots, x_{n+1}), \ldots, p_n(x_2, \ldots, x_{n+1})) \\
q_2 &= (p_n(x_3, \ldots, x_{n+1}, x_1), 0, p_1(x_3, \ldots, x_{n+1}, x_1), \ldots, p_{n-1}(x_3, \ldots, x_1)) \\
&\vdots \\
q_{n+1} &= (p_1(x_1, \ldots, x_n), p_2(x_1, \ldots, x_n), \ldots, p_n(x_1, \ldots, x_n), 0).
\end{aligned}$$

Add these functions and normalize

$$P = \frac{1}{\sqrt{n(n+1)}} \sum_{i=1}^{n+1} q_i,$$

so that $\sum P_i = \sqrt{n+1}$ identically.

This vector field can now be inserted into Theorem 2.2.3, to give the estimate

$$\begin{aligned} J_{n+1} &\leq \int_{\mathbb{R}^{n+1}} \left(\frac{\operatorname{div}(P w_{n+1})}{w_{n+1}} \right)^2 w_{n+1} \\ &= \frac{1}{n(n+1)} \int_{\mathbb{R}^{n+1}} \left(\sum_{i=1}^{n+1} \frac{\operatorname{div}(q_i w_{n+1})}{w_{n+1}} \right)^2 w_{n+1} \end{aligned}$$

Our goal is to estimate this by J_n, which can be written as

$$J_n = \frac{1}{n+1} \int_{\mathbb{R}^{n+1}} \sum_{i=1}^{n+1} \left(\frac{\operatorname{div}(q_i w_{n+1})}{w_{n+1}} \right)^2 w_{n+1}.$$

Let y_i denote $\frac{\operatorname{div}(q_i w_{n+1})}{w_{n+1}}$. Our aim is to show that in $L_2(w_{n+1})$ (the Hilbert space with weight w_{n+1}):

$$\|y_1 + \cdots + y_{n+1}\|^2 \leq n \left(\|y_1\|^2 + \cdots \|y_{n+1}\|^2 \right).$$

With no assumptions on the y_i, the Cauchy-Schwarz inequality would give a coefficient of $(n+1)$ instead of n on the right hand side. However, our y_i have additional special properties, which will ensure this slightly stronger inequality. Define $T_1 : L_2(w) \to L_2(w)$ by

$$(T_1 \phi)(x) = \int \phi(u, x_2, x_3, \ldots, x_n) f(u) du$$

(so that $T_1(\phi)$ is independent of the first coordinate) and similarly for T_i, by integrating out the i^{th} coordinate. Then the T_i are commuting orthogonal projections on the Hilbert space $L_2(w)$. Moreover for each i we have $T_i y_i = y_i$, since y_i is already independent of the i^{th} coordinate. Also, for each j we have $T_1 \cdots T_{n+1} y_j = 0$, because we integrate a divergence. These are the properties which ensure the slightly stronger inequality that we need. We summarize it in the following lemma which completes the proof of Theorem 2.1.1.

Lemma 2.3.1 *Let $T_1, \ldots T_m$ be m commuting orthogonal projections in a Hilbert space H. Assume that we have m vectors $y_1, \ldots, y_m$ such that $T_1 \cdots T_m y_j = 0$ for every $1 \leq j \leq m$. Then*

$$\|T_1 y_1 + \cdots + T_m y_m\|^2 \leq (m-1) \left(\|y_1\|^2 + \cdots \|y_m\|^2 \right).$$

Proof Since the projections are commuting, we can decompose the Hilbert space $H = \bigoplus_{\varepsilon \in \{0,1\}^m}^{\perp} H_\varepsilon$ where $H_\varepsilon = \{x : T_i x = \varepsilon_i x, 1 \leq i \leq m\}$. This is an orthogonal decomposition, so that for $\phi \in H$, $\|\phi\|^2 = \sum_{\varepsilon\in\{0,1\}^m} \|\phi_\varepsilon\|^2$. We decompose each y_i separately as, $y_i = \sum_{\varepsilon\in\{0,1\}^m} y^i_\varepsilon$. The condition in the statement of the lemma implies that $y^i_{(1,\ldots,1)} = 0$ for each i. Therefore

$$T_1y_1 + \cdots + T_my_m = \sum_{i=1}^{m} \sum_{\varepsilon\in\{0,1\}^m} T_iy^i_\varepsilon = \sum_{\varepsilon\in\{0,1\}^m} \sum_{\varepsilon_i=1} y^i_\varepsilon.$$

Now we can compute the norm of the sum as follows

$$\|T_1y_1 + \cdots + T_my_m\|^2 = \sum_{\varepsilon\in\{0,1\}^m} \|\sum_{\varepsilon_i=1} y^i_\varepsilon\|^2.$$

Every vector on the right hand side is a sum of at most $m-1$ summands, since the only vector with m 1's does not contribute anything to the sum. Thus we can complete the proof of the lemma using the Cauchy-Schwarz inequality:

$$\|T_1y_1 + \cdots + T_ny_m\|^2 \leq \sum_{\varepsilon\in\{0,1\}^m} (m-1) \sum_{\varepsilon(i)=1} \|y^i_\varepsilon\|^2 = (m-1)\sum_{i=1}^{m} \|y_i\|^2.$$

□

We end this section with two generalizations. Firstly we point out that the result applies also to the case where X is a random vector. In this case the Fisher information is realized as a matrix, which in the sufficiently smooth case can be written as

$$J_{i,j} = \int \left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right)\frac{1}{f}.$$

Theorem 2.2.3 generalizes in the following way. Let w be a density defined on $\mathbb{R}^n$ and let h be the (vector) marginal of w on the subspace E. For $x \in E$ we define $h(x) = \int_{x+E^\perp} w$. Then for a unit vector e in E

$$\langle Je, e\rangle = \inf \int \left(\frac{\operatorname{div}(wp)}{w}\right)^2 w,$$

where the infimum runs over all $p : \mathbb{R}^n \to \mathbb{R}^n$ for which the orthogonal projection of p into E is constantly e. The rest of the argument is exactly the same as in the one-dimensional case.

Our second remark concerns the assumption that the variables are identically distributed. In the case of identically distributed random variables one may invoke the symmetry of the situation to prove the existence of a minimizing p in Theorem 2.2.3 with some additional properties. Thus one may actually choose p so that for every i, $p_i(x_1, \ldots, x_n) = p_1(x_i, \ldots, x_{i-1})$, and also that the value of $p_1(x_1, \dots, x_n)$ does not depend on the order of the arguments, beyond the first, $x_2, \dots, x_n$. However, in the proof we did not use any of these properties. Thus, in fact the proof yields a more general result about Fisher information of the sum of independent random variables

Theorem 2.3.2 *For any independent random variables $X_1, \dots X_{n+1}$ with variance 1 and sufficiently regular densities,*

$$J\left(\frac{1}{\sqrt{n+1}}\sum_{i=1}^{n+1} X_i\right) \leq \frac{1}{n+1}\sum_{j=1}^{n+1} J\left(\frac{1}{\sqrt{n}}\sum_{i\neq j} X_i\right).$$

This inequality can be integrated along the Ornstein-Uhlenbeck semigroup, as described in Section 2.2.1, to give the non-identically distributed version of Theorem 2.1.1:

Theorem 2.3.3 *Let $X_1, X_2, \dots$ be independent random variables with variance 1. Then*

$$\mathrm{Ent}\left(\frac{X_1 + \cdots + X_{n+1}}{\sqrt{n+1}}\right) \geq \frac{1}{n+1}\sum_{j=1}^{n+1} \mathrm{Ent}\left(\frac{1}{\sqrt{n}}\sum_{i\neq j} X_i\right).$$

2.3.1 The Entropy Power Inequality

In this section we prove the following

Theorem 2.3.4 *Let X_i, $i = 1, \dots, n$ be independent random variables which satisfy the same regularity assumptions as in Theorem 2.1.1. Then*

$$e^{2\mathrm{Ent}\sum_{i=1}^n X_i} \geq \frac{1}{n-1}\sum_{j=1}^n e^{2\mathrm{Ent}\sum_{i\neq j}^n X_i}. \tag{2.3.1}$$

This Theorem is more general than Theorem 2.1.1. Indeed, if we assume that X_i are i.i.d., and normalize properly, then equation (2.3.1) reads

$$e^{2(\mathrm{Ent}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i + \ln\sqrt{n})} \geq \frac{1}{n-1}\sum_{j=1}^n e^{2(\mathrm{Ent}\frac{1}{\sqrt{n-1}}\sum_{i\neq j}^n X_i + \ln\sqrt{n-1})},$$

and we can rewrite it as

$$\mathrm{Ent}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \geq \frac{1}{2}\ln[\frac{1}{n}\sum_{j=1}^n e^{2\mathrm{Ent}\frac{1}{\sqrt{n-1}}\sum_{i\neq j}^n X_i}].$$

Use the concavity of the logarithmic function to get

$$\operatorname{Ent}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) \geq \frac{1}{n}\sum_{j=1}^{n}\operatorname{Ent}\left(\frac{1}{\sqrt{n-1}}\sum_{i\neq j}^{n} X_i\right).$$

Since the variables were i.i.d., we get Theorem 2.1.1. Moreover, we can actually get from here the not identically distributed case as well by similar reasoning.

The proof of this theorem is in the same lines as of Theorem 2.1.1, only with some more complications because one has to consider the more general case. We include it below with all the details.

The first step is to get an estimate for the information of the sum $\sum_{i=1}^{n} a_i X_i$ in terms of informations of the partial (and normalized) sums $\sum_{i\neq j}^{n} a_i X_i / \sqrt{1-a_j^2}$, where we assume that $\mathrm{Var}(X_i) = 1$. We prove the following lemma

Lemma 2.3.5 *For independent square integrable random variables X_i, $i = 1, \ldots n$, and a unit vector $\hat{a} = (a_1, \ldots a_n)$,*

$$J\left(\sum_{i=1}^{n} a_i X_i\right) \leq \sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1} J\left(\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j}^{n} a_i X_i\right)$$

Proof Denote $\hat{a}_j = (a_1, \ldots, a_{j-1}, 0, a_{j+1} \ldots, a_n)/(1-a_j^2)$, it is a unit vector, and can also be viewed as a (unit) vector in R^{n-1}. Denote

$$w(x_1, \ldots, x_n) = f_1(x_1)\cdots f_n(x_n),$$

where f_i is the density of X_i. Assume for every j that the function

$$p_j = (p_1^j, \ldots, p_{j-1}^j, 0, p_{j+1}^j, \ldots, p_n^j)$$

is the function guaranteed from Theorem 2.2.3, which realizes the information for $\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j}^{n} a_i X_i$, i.e.,

$$J\left(\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j}^{n} a_i X_i\right) = \int \left(\frac{\mathrm{div} w p_j}{w}\right)^2 w,$$

and $\langle p_j, \hat{a}\rangle = 1$. We construct an n-dimensional function p with $\langle p, \hat{a}\rangle = 1$ in the following way (note that p_j were, essentially, $(n-1)$-dimensional)

$$p = \sum_{j=1}^{n} b_j p_j.$$

The condition on the b_j's is that $\sum b_j\sqrt{1-a_j^2} = 1$. In particular, we may choose $b_j = \sqrt{1-a_j^2}/(n-1)$.

We use Theorem 2.2.3 again, this time to bound

$$J\left(\sum_{i=1}^{n} a_i X_i\right) \leq \int \left(\frac{\mathrm{div} wp}{w}\right)^2 w.$$

We insert the definition of p, the choice of p_j, and, most important, we use the Hilbert Space Lemma 2.3.1. In applying the lemma we use the fact that each of the p_j's is independent of the j^{th} coordinate. We thus arrive at the following string of inequalities

$$J\left(\sum_{i=1}^{n} a_i X_i\right) \leq \int \left(\sum_{j=1}^{n} b_j \frac{\mathrm{div} wp_j}{w}\right)^2 w$$

$$\leq (n-1)\sum_{j=1}^{n} b_j^2 \int \left(\frac{\mathrm{div} wp_j}{w}\right)^2 w = (n-1)\sum_{j=1}^{n} b_j^2 J\left(\sum_{i\neq j}^{n} a_i X_i/\sqrt{1-a_j^2}\right).$$

For our choice of b_j we can rewrite this in the form

$$J\left(\sum_{i=1}^{n} a_i X_i\right) \leq \sum_{j=1}^{n} \frac{1-a_j^2}{n-1} J\left(\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j} a_i X_i\right).$$

□

Corollary 2.3.6 *For independent square integrable random variables X_i, $i = 1,\dots n$, and a unit vector $\hat{a} = (a_1,\dots a_n)$,*

$$\mathrm{Ent}\left(\sum_{i=1}^{n} a_i X_i\right) \geq \sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1} \mathrm{Ent} \frac{1}{\sqrt{1-a_j^2}} \left(\sum_{i\neq j}^{n} a_i X_i\right) \tag{2.3.2}$$

Proof As described in Section 2.2.1, we may integrate along the Ornstein Uhlenbeck adjoint semigroup. We need to be careful however, since now our random variables do not have equal variance. Denote the variance of X_i by σ_i^2. Subtract for each side of the information inequality the quantity $\sum a_i^2\sigma_i^2$, which factors in the right hand side to $\sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1} \frac{\sum_{i\neq j}^{n} a_i^2\sigma_i^2}{1-a_j^2}$. Do this separately at each time t, and integrate along the semigroup. We get the following inequality

$$\mathrm{Ent}\left(\sum a_i^2\sigma_i^2 G\right) - \mathrm{Ent}\left(\sum_{i=1}^{n} a_i X_i\right) \leq$$

$$\sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1}\left(\mathrm{Ent}\left(\frac{\sum_{i\neq j}^{n} a_i^2\sigma_i^2}{1-a_j^2} G\right) - \mathrm{Ent}\left(\frac{1}{\sqrt{1-a_j^2}}\sum_{i\neq j}^{n} a_i X_i\right)\right),$$

which can be rewritten as (rearranging, and cancelling the terms $\mathrm{Ent}(G)$ on both sides)

$$\mathrm{Ent}\left(\sum_{i=1}^{n} a_i X_i\right) \geq \sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1} \mathrm{Ent}\left(\frac{1}{\sqrt{1-a_j^2}} \sum_{i \neq j}^{n} a_i X_i\right) +$$
$$+ \log\left(\sum a_i^2 \sigma_i^2\right) - \sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1} \log\left(\frac{1}{\sqrt{1-a_j^2}} \sum_{i \neq j}^{n} a_i X_i\right).$$

Using the concavity of the logarithmic function, and the fact that $\sum_{j=1}^{n} \frac{(1-a_j^2)}{n-1} = 1$, the proof of the corollary is complete. $\square$

Proof of Theorem 2.3.4 For every $(b_1, \ldots, b_n)$ with $\sum b_i^2 = 1$, Corollary 2.3.6 implies that

$$\mathrm{Ent} \sum_{i=1}^{n} X_i = \mathrm{Ent} \sum_{i=1}^{n} b_i \left(\frac{X_i}{b_i}\right) \geq \frac{1}{n-1} \sum_{j=1}^{n} (1-b_j^2) \mathrm{Ent}\left(\frac{1}{1-b_j^2} \sum_{i \neq j}^{n} X_i\right).$$

Introducing the new parameter $c_i = 1 - b_i^2$, and denoting $\alpha_j = \mathrm{Ent}\left(\sum_{i \neq j} X_i\right)$, we have that

$$\mathrm{Ent} \sum_{i=1}^{n} X_i \geq \frac{1}{n-1} \sum_{j=1}^{n} c_j \alpha_j - \frac{1}{2} c_j \log c_j,$$

for every $0 \leq c_j \leq 1$ with $\sum_{j=1}^{n} c_j = n-1$. We optimize this estimate over all admissible sequences $(c_j)_{j=1}^{n}$. We use a Lagrange multiplier λ,

$$\alpha_j - \frac{1}{2}(\log c_j + 1) = \lambda, \qquad \sum_{j=1}^{n} c_j = n-1;$$
$$c_j = e^{2\alpha_j - 2\lambda - 1}, \qquad \sum_{j=1}^{n} e^{2\alpha_j - 2\lambda - 1} = n-1,$$

and conclude that $\lambda = \frac{1}{2} \log \frac{1}{n-1} \sum_{j=1}^{n} e^{2\alpha_j - 2\lambda - 1}$ and $c_j = (n-1) \frac{e^{2\alpha_j - 2\lambda - 1}}{\sum e^{2\alpha_i - 2\lambda - 1}}$. To see what bound we obtained, insert this in to the formula above,

$$\frac{1}{n-1}\left(\sum_{j=1}^{n} c_j (\alpha_j - \frac{1}{2}\log c_j)\right) = \frac{1}{n-1}\left(\sum_{j=1}^{n} e^{2\alpha_j - 2\lambda - 1}(\alpha_j - \frac{1}{2}(2\alpha_j - 2\lambda - 1))\right)$$
$$= \frac{1}{n-1}\left(\sum_{j=1}^{n} (2\lambda+1) e^{2\alpha_j - 2\lambda - 1}\right) = \frac{1}{n-1}(2\lambda+1) e^{-2\lambda} \sum_{j=1}^{n} e^{2\alpha_j - 2\lambda - 1} = 2\lambda + 1.$$

In other words,

$$e^{2\mathrm{Ent}\sum_{i=1}^n X_i} \geq \frac{1}{n-1}\sum_{j=1}^n e^{2\mathrm{Ent}\sum_{i\neq j}^n X_i}.$$

This completes the demonstration of the entropy power result. (The scale invariance of the entropy power result allows us to take any positive numbers $a_1, \ldots, a_n$, which in turn imply the result for independent random variables with arbitrary variances.) □

Theorem 2.3.1 is a generalization of Shannon's entropy power inequality (see [Lie])

$$e^{2\mathrm{Ent}(X+Y)} \geq e^{2\mathrm{Ent}(X)} + e^{2\mathrm{Ent}(Y)}.$$

Remark To avoid Lagrange multipliers one may simply take inequality (2.3.2) with $\lambda_j = \frac{1-a_j^2}{n-1}$ and $Y_i = a_i X_i$. It reads (for $\sum \lambda_j = 1$),

$$\mathrm{Ent}\sum_{i=1}^n Y_i \geq \sum_{j=1}^n \lambda_j \mathrm{Ent}\frac{1}{\sqrt{(n-1)\lambda_j}}\sum_{i\neq j}^n Y_i.$$

Taking e^2 to this power, we get

$$e^{2\mathrm{Ent}\sum_{i=1}^n Y_i} \geq \Pi_j \left(e^{2\mathrm{Ent}\frac{1}{\sqrt{(n-1)\lambda_j}}\sum_{i\neq j}^n X_i}\right)^{\lambda_j} =$$

$$\Pi_{j=1}^n \left(e^{2\mathrm{Ent}\left(\sum_{i\neq j}^n Y_i\right)-\log(n-1)\lambda_j}\right)^{\lambda_j} = \Pi_{j=1}^n \left(\frac{\frac{1}{n-1}e^{2\mathrm{Ent}\left(\sum_{i\neq j}^n Y_i\right)}}{\lambda_j}\right)^{\lambda_j}.$$

This holds for every choice of λ_j, $1 \leq \lambda \leq n$, satisfying $\sum \lambda_j = 1$. We can optimize over the choice of λ_j's (choosing $\lambda_j = \frac{\alpha_j}{\sum \alpha_i}$ for $\alpha_j = \frac{1}{n-1}e^{2\mathrm{Ent}\left(\sum_{i\neq j}^n Y_i\right)}$), or use the fact that we know that the maximum over all such geometric means is the arithmetic mean, to establish the final result

$$e^{2\mathrm{Ent}\sum_{i=1}^n Y_i} \geq \frac{1}{n-1}\sum_{j=1}^n e^{2\mathrm{Ent}\left(\sum_{i\neq j}^n Y_i\right)}.$$

2.4 Proof of Quantitative Entropic CLT

In this section we will prove Theorem 2.1.3. The considerations of Section 2.2.1 have already reduced this to the proof of Theorem 2.2.2.

Let X with density f be as in Theorem 2.2.2, and assume as we may that the variable is centered, that is $\mathbb{E}X = \int xf(x)\,dx = 0$. Set, as before, $w(x_1, \ldots, x_n) = f(x_1)\cdots f(x_n)$. Then the function $h(t) = \int_{ta+a^\perp} w$ is the density of $\sum_{i=1}^n a_i X_i$. The variational representation of Fisher information of marginals, in the form of Lemma 2.2.4 yields

$$J\left(\sum_{i=1}^n a_i X_i\right) \leq \int w\left[Tr(Dp)^2 + \langle \mathrm{Hess}(-\log w)p, p\rangle\right], \tag{2.4.1}$$

for any sufficiently integrable vector field $p : \mathbb{R}^n \to \mathbb{R}^n$ with the property that for every x, $\langle p(x), a\rangle = 1$. Note that the best choice of p would give equality in (2.4.1), whereas the simplest choice ($p(x) = a$ for all x) recovers the result $J\left(\sum_{i=1}^n a_i X_i\right) \leq J(X)$, which is a consequence of the Blachman-Stam inequality [Bla, Sta]. Our strategy will be to optimize over a workable sub-family of test functions p.

2.4.1 Choice of particular test functions

In order to exploit the product structure of w, in expression (2.4.1), we restrict ourselves to functions p_i which are linear combinations of the constant function and functions depending only on one variable. It seems to be sufficient to choose a single one-variable function: so our test function has the form $p_i(x_1, \ldots, x_n) = c_{i0} + \sum_{j=1}^n c_{ij} r(x_j)$ and the only restriction on p is that $\sum_{i=1}^n a_i p_i \equiv 1$. We would like to choose $p = (a_1 r(x_1), a_2 r(x_2), \ldots, a_n r(x_n))$ but since this does not satisfy the condition we project it orthogonally onto the hyperplane $\{x : \langle a, x\rangle = 1\}$. So we choose

$$p_i(x_1, \ldots, x_n) = a_i + a_i(1 - a_i^2) r(x_i) - a_i \sum_{j \neq i} a_j^2 r(x_j),$$

where r is a real-valued function on $\mathbb{R}$.

Now for smooth r with compact support inequality (2.4.1) is valid. Denoting $\phi = -\log f$, the two integrands on the right hand side of (2.4.1) are

$$\mathrm{Tr}\Big((Dp)^2\Big) = \sum_{1 \leq i,j \leq n} \frac{\partial p_i}{\partial x_j}\frac{\partial p_j}{\partial x_i} = \sum_{i=1}^n a_i^2(1 - a_i^2)^2 r'(x_i)^2 + \sum_{i \neq j} a_i^3 a_j^3 r'(x_i) r'(x_j),$$

and

$$\begin{aligned}
\langle \text{Hess}(-\log w)p, p\rangle &= \sum_{i=1}^{n} \phi''(x_i)a_i^2 \left(1 + (1-a_i^2)r(x_i) - \sum_{j\neq i} a_j^2 r(x_j)\right)^2 \\
&= \sum_{i=1}^{n} \phi''(x_i)a_i^2 \Bigg[1 + (1-a_i^2)^2 r(x_i)^2 + \sum_{j\neq i} a_j^4 r(x_j)^2 - 2\sum_{j\neq i} a_j^2 r(x_j) \\
&\qquad +2(1-a_i^2)r(x_i) - 2\sum_{j\neq i}(1-a_i^2)a_j^2 r(x_j)r(x_i) + \\
&\qquad \sum_{i\neq j\neq k\neq i} a_j^2 a_k^2 r(x_j)r(x_k)\Bigg].
\end{aligned}$$

Integrating against the product density $w(x)$ and using the fact that $\sum a_i^2 = 1$, this identity reduces to

$$\begin{aligned}
&\int w\left[Tr(Dp)^2 + \langle \text{Hess}(-\log w)p, p\rangle\right] \\
= \quad & J + W\left(\int f(r')^2 + \int f\phi'' r^2\right) + JV\int fr^2 + J(W-V)\left(\int fr\right)^2 \\
& +2U\left(\int f\phi'' r - J\int fr\right) - 2W\left(\int fr\right)\left(\int f\phi'' r\right) + M\left(\int fr'\right)^2,
\end{aligned}$$

where:

$$J = J(X), \quad U = 1 - \sum_{i=1}^{n} a_i^4, \quad V = \sum_{i=1}^{n} a_i^4 - \sum_{i=1}^{n} a_i^6, \quad W = U - V$$
$$\text{and } M = \left(\sum_{i=1}^{n} a_i^3\right)^2 - \sum_{i=1}^{n} a_i^6.$$

Our aim is to choose r to make the estimate small. This is a standard variational problem except for the term $M\left(\int fr'\right)^2$. If the random variable X is symmetric, so that f is an even function, then we will choose an even function r and the term in question will vanish. This gives a slightly better result for symmetric random variables. In the the general case we can control the term because

$$\left(\int fr'\right)^2 = \left(\int f'r\right)^2 \leq \left(\int \frac{(f')^2}{f}\right)\left(\int fr^2\right) = J\left(\int fr^2\right). \tag{2.4.2}$$

Our choice of a vector field p is invariant under the addition of a constant to r and this property simplifies considerably the calculations. In order to maintain this invariance we improve the estimate (2.4.2) slightly as follows. Setting $m = \int fr$ we have

$$\left(\int fr'\right)^2 = \left(\int f'(r-m)\right)^2 \leq \left(\int \frac{(f')^2}{f}\right)\left(\int f(r-m)^2\right) = J\left(\int fr^2 - m^2\right). \tag{2.4.3}$$

Lastly, the coefficient M is at most V by Hölder's inequality.

Hence we obtain the estimate

$$J\left(\sum_{i=1}^{n} a_i X_i\right) - J \leq T(r)$$

where

$$\begin{aligned} T(r) &= W\left(\int f(r')^2 + \int f\phi'' r^2\right) + 2JV\int fr^2 + J(W-2V)\left(\int fr\right)^2 \\ &\quad +2U\left(\int f\phi'' r - J\int fr\right) - 2W\left(\int fr\right)\left(\int f\phi'' r\right). \end{aligned}$$

Our goal is to find r that makes this quantity as negative as possible.

2.4.2 Optimization in the test function r

In this subsection we apply variational methods to find the optimum value of $T(r)$. Their statement in the appropriate Sobolev spaces, as well as the questions of existence and regularity are discussed in detail in the appendix of [BBN]. So, in the following we shall ignore technical issues.

In order to deal with the terms $J(W-2V)\left(\int fr\right)^2$ and $-2W\left(\int fr\right)\left(\int f\phi'' r\right)$ we introduce a Lagrange multiplier λ and choose r to be λr_0 where r_0 is the minimizer of

$$\int f(r')^2 + \int f\phi'' r^2 + J\beta\int fr^2 + 2\left(\int f\phi'' r - J\int fr\right)$$

and $\beta = 2V/W$. The Euler-Lagrange equation for this minimization problem is

$$-(fr_0')' + fr_0\phi'' + J\beta fr_0 = Jf - f\phi''. \tag{2.4.4}$$

By integrating this equation against r_0 and 1 respectively, we obtain that any solution r_0 of (2.4.4) automatically satisfies the following relations

$$\int f(r_0')^2 + \int f\phi'' r_0^2 + J\beta\int fr_0^2 = J\int fr_0 - \int f\phi'' r_0, \tag{2.4.5}$$

$$\int f\phi'' r_0 + J\beta\int fr_0 = 0. \tag{2.4.6}$$

Set $A = \int fr_0$ and $B = \int f\phi'' r_0$. Using equations (2.4.5) and (2.4.6) we obtain

$$T(\lambda r) = AJ\lambda(1+\beta)\left(\lambda(1+A)W - 2U\right).$$

The optimal choice of λ is $U/W(1+A)$ and the estimate is then

$$T(\lambda r_0) = -\frac{JU^2(1+\beta)}{W}\frac{A}{A+1}. \tag{2.4.7}$$

To finish the proof we need to estimate $A = \int fr_0$ from below: we want to know that if U and W are close to 1 and β is small then A is large so that the quantity above captures almost all of J. From now on, we suppress the subscript and use r instead of r_0. The equation satisfied by r reads

$$-(fr')' + f\phi'' r + J\beta fr = f(J - \phi''). \tag{2.4.8}$$

Writing $f' = -f\phi'$ and dividing by f we can rewrite it as

$$-r'' + \phi' r' + \phi'' r + J\beta r = J - \phi''.$$

If we set $r = s'$ then we can integrate the equation to get

$$-s'' + \phi' s' + J\beta s = Jx - \phi' \tag{2.4.9}$$

where the constant of integration is chosen so that $\int fs = 0$ (assuming, as we are, that our random variable is centered so that $\int fx = 0$). Multiplying (2.4.9) by f we get

$$-\left(fs'\right)' + J\beta fs = Jfx - f\phi' \tag{2.4.10}$$

Now equation (2.4.10) is the Euler-Lagrange equation for a different minimization problem in which the objective function is

$$Q(s) = \frac{1}{2}\int f(s')^2 + \frac{J\beta}{2}\int f\left(s - \frac{x}{\beta} + \frac{\phi'}{J\beta}\right)^2. \tag{2.4.11}$$

Observe that in our new notation $A = \int fs'$. Integrating equation (2.4.10) against s and x respectively we get

$$\int f(s')^2 + J\beta\int fs^2 = J\int fsx - \int fs'$$

and

$$\int fs' + J\beta\int fsx = J\int fx^2 - \int f = J - 1$$

since f is the density of a random variable with variance 1. As before these equations yield an expression of $Q(s)$ in terms of $A = \int fs'$ only. One obtains that

$$Q(s) = \frac{1+\beta}{2\beta}A. \tag{2.4.12}$$

In order to bound A from below it thus suffices to bound the functional Q. Note that the value of Q on the zero function is

$$Q(0) = \frac{J}{2\beta}\int f\left(x - \frac{\phi'}{J}\right)^2 = \frac{J-1}{2\beta}.$$

This is where we use the fact that f satisfies a Poincaré inequality with constant c. This will force the minimum of Q to be of the order of $Q(0)$. Since $\int fs = 0$ the Poincaré inequality (2.1.3) gives

$$c\int fs^2 \le \int fs'^2.$$

Thus we get

$$\begin{aligned} Q(s) &= \frac{1}{2}\int fs'^2 + \frac{J\beta}{2}\int f\left(s - \frac{x}{\beta} + \frac{\phi'}{J\beta}\right)^2 \\ &\geq \frac{1}{2}\int f\left(cs^2 + J\beta\left(s - \frac{x}{\beta} + \frac{\phi'}{J\beta}\right)^2\right) \\ &\geq \frac{1}{2}\cdot\frac{cJ\beta}{c+J\beta}\int f\left(-\frac{x}{\beta} + \frac{\phi'}{J\beta}\right)^2 \\ &= \frac{c}{2(c+J\beta)}Q(0) = \frac{c(J-1)}{2\beta(c+\beta J)}. \end{aligned}$$

Substituting into the expression (2.4.12) for $Q(s)$ in terms of A, we obtain the lower bound

$$A \ge \frac{c(J-1)}{(1+\beta)(c+J\beta)}.$$

Substituting this into (2.4.7), rearranging, and using the facts that $W = U - V$ and $\beta = 2V/W$, gives an estimate slightly better than

$$J\left(\sum_{i=1}^{n} a_i X_i\right) - J \le -\frac{c(J-1)U^2}{cU + 2V - cV}, \qquad (2.4.13)$$

where we have used the fact that $J \ge 1$, which is a consequence of the Cramér-Rao inequality. It is easy to check that $V \le U - U^2$ and this implies the desired estimate

$$J\left(\sum a_i X_i\right) - 1 \le (J-1)\frac{1-U}{cU/2 + 1 - U} = (J-1)\frac{\alpha(a)}{c(1-\alpha(a))/2 + \alpha(a)}.$$

$\square$

Chapter 3

Applications and related topics

3.1 Introduction for the Chapter

In this chapter we describe further results either concerning entropy and covering or using them as tools. In the first section we establish the concentration, as n tends to ∞, of the Haar measure on the sphere S^{n-1}, in distance $\sqrt{1-\lambda}$ from λn-dimensional subspaces. We provide precise asymptotics for the concentration rates. These asymptotics are applied to a problem of projections of points on the sphere, and to confirming a one sided estimate of the duality of entropy numbers conjecture, in a special case. (This confirmation is a partial case of the theorems proved in Chapter 1, but the method is of independent interest.)

In the second section we provide an estimate for the diameter of a sign-projection of a convex body. To do this we apply a strong theorem of M. Talagrand called the Majorizing Measure Theorem (see [Ta1]), which establishes the existence of a special family of nets in different resolutions on a convex body.

In the third section we have gathered some further remarks concerning covering numbers. Firstly, we examine the behavior of the function $e_k(P_tK)$, the k^{th} entropy number of a random rank t projection of the convex body K. We show, among other things, that it exhibits in some cases a phase transition at a certain level k depending on the parameters of the convex body. Secondly, we define the notion of weighted covering number, which is a natural extension of the standard covering (and exists already in combinatorics). We show that in our context this notion is equivalent to the standard covering number up to universal constants. In the last page we remark about covering which allows rotations, and illustrate that in that case a duality relation cannot hold.

This chapter includes results from the papers [Art1], [Art2] and [AMS3].

3.2 Proportional Concentration on the Sphere

3.2.1 Introduction

In this section we establish concentration phenomena for subspaces with arbitrary dimension. Namely, we display conditions under which the Haar measure on the sphere concentrates on a neighborhood of the intersection of the sphere with a subspace of $\mathbb{R}^n$ of a given dimension. We display applications to a problem of projections of points on the sphere, and to the duality of entropy numbers conjecture.

The classical concentration phenomenon refers to concentration of the Haar measure on the sphere S^{n-1} around the intersection with $(n-1)$-dimensional subspaces. Namely, for any fixed ε, the area in $S^{n-1}$of the ε-neighborhood of an $(n-1)$-dimensional subspace, converges to 1 when $n \to \infty$. This phenomenon and its extensions have attracted much attention, starting from the paper [Mi1], and have a variety of applications in the Asymptotic Theory of Normed Spaces, see e.g. [MSch], [GiMi]. The complementary case is that of the area in S^{n-1} of the ε-neighborhood of a 1-dimensional subspace. It converges, of course, to 0 when $n \to \infty$. In both, the case of dimension 1 and the case of co-dimension 1, the convergence is exponentially fast, a fact which plays a central role in the applications.

The results displayed here concern the limit as $n \to \infty$ of the area in S^{n-1} of the ε-neighborhood of a k-dimensional subspace E_k, for the full range of k between 3 and $n-3$. A particular case is when k is proportional to n, namely, $k = \lambda n$. In this case it is not difficult to show (we do this in Section 3.2.2) that for every $0 < \lambda < 1$, there exists a critical value $\varepsilon(\lambda)$, determined by the formula $\sin \varepsilon(\lambda) = \sqrt{1-\lambda}$, such that: If $\varepsilon > \varepsilon(\lambda)$ then $\mu((E_k)_\varepsilon) \to 1$ as $n \to \infty$, and if $\varepsilon < \varepsilon(\lambda)$, then $\mu((E_k)_\varepsilon) \to 0$ as $n \to \infty$ (as customary, $(E_k)_\varepsilon$ denotes the set $\{x \in S^{n-1} : \rho(x, E_k \cap S^{n-1}) < \varepsilon\}$ where ρ is the geodesic distance on the sphere, and μ denotes the normalized rotation invariant Haar measure on the sphere). Our main concern is finding the exact rates of convergence. Approximations for these rates turn out to be useful in several applications. We show that in both, the case $\varepsilon > \varepsilon(\lambda)$, and the case $\varepsilon < \varepsilon(\lambda)$, the rate of convergence is exponential in n, namely $1 - e^{-\gamma n}$ and $e^{-\gamma n}$ respectively, where

$\gamma = \gamma(\lambda, \delta)$ is a constant depending only on λ and on $\delta = \varepsilon - \varepsilon(\lambda)$. We provide precise estimates for the constant $\gamma(\lambda, \delta)$. This is done in Sections 3.2.3, 3.2.4 and 3.2.5.

The established estimates are then applied to several problems in the Asymptotic Theory of Normed Spaces. The first application, in Section 3.2.6, deals with projections of points on the sphere into lower dimensional subspaces as follows. A random point on S^{n-1}, projected on a λn-dimensional subspace, has with high probability euclidean norm close to $\sqrt{\lambda}$. We estimate precisely this probability. This enables us, for instance, to provide an isomorphic Johnson-Lindenstrauss lemma. The second application, in Section 3.2.7, is concerned with the duality of entropy numbers conjecture. In the special case where one of the bodies is the euclidean ball D_n, and where the covering number of K (a general convex body) by D_n is exponentially large, we show duality, namely, that the covering number of D_n by the polar body K° is also exponentially large. By now the conjecture, in the case where one of the bodies is D_n, has been proved in full generality (see Chapter 1); however, in the special case described here, with use of our estimates it is just a a straightforward computation. In Section 3.2.8 we give another application of the estimates to compare, following Diaconis and Freedman, the distribution of the first k coordinates of a random vector on the $(n-1)$-dimensional sphere with a random gaussian vector with k coordinates. Finally, in Section 3.2.9 we remark on a very beautiful recent result by M. Gromov [Gr] and its relevance to the estimates we have established.

3.2.2 A useful representation of the spherical area

In this section we represent the area of the ε-neighborhood of a k-dimensional subspace by means of Beta-distributed random variables. Through this representation, some of the properties of the behavior of the area become transparent. The area, on the sphere, of the ε-neighborhood of a k-dimensional subspace is

$$\mu((E_k)_\varepsilon) = \frac{1}{\int_0^{\pi/2} \sin^{n-k-1} x \cos^{k-1} x dx} \int_0^\varepsilon \sin^{n-k-1} x \cos^{k-1} x dx.$$

By the change of variables $u = \sin^2 x$ one gets

$$\mu((E_k)_\varepsilon) = \frac{1}{\int_0^1 u^{\frac{n-k}{2}-1}(1-u)^{\frac{k}{2}-1} du} \int_0^{\sin^2 \varepsilon} u^{\frac{n-k}{2}-1}(1-u)^{\frac{k}{2}-1} du. \qquad (3.2.1)$$

Recall that the integral $\int_0^1 u^{m-1}(1-u)^{l-1} du$ coincides with the Beta function $\text{Beta}(m, l) = \frac{\Gamma(m)\Gamma(l)}{\Gamma(m+l)} = \frac{(m-1)!(l-1)!}{(m+l-1)!}$. This can be easily verified by induction.

The expression (3.2.1) implies, therefore, that as a function of ε, the measure $\mu((E_k)_\varepsilon)$ is the distribution of a Beta random variable with parameters $(\frac{n-k}{2}, \frac{k}{2})$. The expectation of such a variable is $\frac{n-k}{n}$. Letting $k = \lambda n$, we can write

$$\mu((E_k)_\varepsilon) = \frac{\Gamma(n/2)}{\Gamma(\lambda n/2)\Gamma((1-\lambda)n/2)} \int_0^{\sin^2 \varepsilon} u^{\frac{(1-\lambda)n}{2}-1}(1-u)^{\frac{\lambda n}{2}-1} du. \quad (3.2.2)$$

For a random variable Y_n with distribution Beta$((1-\lambda)\frac{n}{2}, \lambda\frac{n}{2})$, we have

$$\begin{aligned} \mu((E_k)_\varepsilon) &= \text{Prob}\,[Y_n \leq \sin^2 \varepsilon], \\ \mathbb{E}[Y_n] &= (1-\lambda), \end{aligned}$$

and

$$\mathrm{Var[Y_n]} = \frac{\lambda(1-\lambda)}{\mathrm{n}/2+1} \to_{\mathrm{n}\to\infty} 0.$$

Remark From the derived representation it is possible to get, easily, the existence of, and the formula for, the critical value $\varepsilon(\lambda)$. The expectation of the distance of a random point x from the subspace E_k is determined by the expectation of Y_n which is independent of n, and it is therefore $\varepsilon(\lambda) = \arcsin(\sqrt{1-\lambda})$. Approximately, half of the measure on the sphere is within distance smaller than $\varepsilon(\lambda)$ from the subspace, since the median and the expectation are close. This already verifies the existence of the critical value of ε. Moreover, the Beta random variables Y_n concentrate, when $n \to \infty$, around their mutual expectation. This holds, for example, since the variance of Y_n tends to 0 as $n \to \infty$. By (3.2.3) we see that, therefore, the measure on the sphere concentrates within distance $\varepsilon(\lambda)$ of the subspace. The main concern of the following sections is to show that this concentration is exponentially fast, and to determine the precise constants.

Remark Note that the ε-neighborhoods we took were with respect to the geodesic distance on the Sphere. If, instead, we would compute the measure with respect to the euclidean distance in $\mathbb{R}^n$, the term "$\sin\varepsilon$" in formula (3.2.2) would become simply "ε".

Finding the critical value was straightforward. A deeper investigation of the asymptotic deviation of a Beta variable from its mean (using methods to handle Beta random variables with large parameters with a constant ratio), is needed in order to get the convergences rates. This is described in the next section.

3.2.3 The main estimate

In this section we display and prove the main result of the paper. Throughout what follows we use the following notation. For variables A and B depending on n, we write $A \simeq B$ in the following two meanings.

1. If both A and B are close to 0, $A \simeq B$ means that $\frac{A}{B} \to 1$ as $n \to \infty$.

2. If both A and B are close to 1, $A \simeq B$ means that $\frac{1-A}{1-B} \to 1$ as $n \to \infty$.

When A and B depend on several parameters, we specify which parameters are fixed, and we mean that the convergence is uniform with respect to all other parameters.

Theorem 3.2.1 *Let E_k be a k-dimensional subspace of $\mathbb{R}^n$, and denote by $\mu((E_k)_\varepsilon)$ the Haar measure on the sphere S^{n-1} of the set of points within geodesic distance smaller than ε of E_k. We write $k = \lambda n$. Fix $0 < \varepsilon < \pi/2$ and $0 < \lambda < 1$. The following estimates hold as $n \to \infty$.*

(i) If $\sin^2 \varepsilon > 1 - \lambda$, then

$$\mu((E_k)_\varepsilon) \simeq 1 - \frac{1}{\sqrt{n\pi}} \frac{\sqrt{\lambda(1-\lambda)}}{\sin^2 \varepsilon - (1-\lambda)} e^{-\frac{n}{2}u(\lambda,\varepsilon)}. \tag{3.2.3}$$

(ii) If $\sin^2 \varepsilon < 1 - \lambda$, then

$$\mu((E_k)_\varepsilon) \simeq \frac{1}{\sqrt{n\pi}} \frac{\sqrt{\lambda(1-\lambda)}}{(1-\lambda) - \sin^2 \varepsilon} e^{-\frac{n}{2}u(\lambda,\varepsilon)} \tag{3.2.4}$$

where $u(\lambda, \varepsilon) = (1-\lambda)\ln \frac{(1-\lambda)}{\sin^2 \varepsilon} + \lambda \ln \frac{\lambda}{\cos^2 \varepsilon}$.

For the proof of Theorem 3.2.1 we need the following preliminary observations. We employ ideas given in Alfers and Dinjes [AlDi], for comparing a Beta$(\alpha m, (1-\alpha)m)$ distributed random variable, with a standard Gaussian random variable, where m is large and $0 < \alpha < 1$ is constant. Following [AlDi], define the two functions:

$$A(\alpha, p) = sign(\alpha - p)\sqrt{2}\left((1-\alpha)\ln(\frac{1-\alpha}{1-p}) + \alpha \ln \frac{\alpha}{p}\right)^{1/2},$$

(thus $u(\lambda, \varepsilon) = A^2((1-\lambda), \sin^2 \varepsilon)/2)$, and

$$D(\alpha, a) = \frac{\sqrt{\alpha(1-\alpha)}}{\alpha - p} A(\alpha, p) \quad \text{where} \quad a = A(\alpha, p).$$

The mapping $A(\alpha,p)$ is well defined (the term in the parentheses is always positive), and for every fixed α it maps $(0,1)$ bijectively onto $(-\infty,\infty)$. Also

$$\frac{\partial}{\partial p}A(\alpha,p) = -\frac{\alpha - p}{p(1-p)A(\alpha,p)}.$$

Notice, however, that if $A(\alpha,p)=0$ then $\alpha = p$, and then $D(\alpha,0)$ is not defined. We define it in the natural way as the limit of $D(\alpha,a)$ when $a\to 0$ and get that $D(\alpha,0)=1$ for every α. A straightforward calculation (using the change of variables $a=A(\alpha,t)$, the observation that $\frac{\partial}{\partial p}A(\alpha,p) = -\frac{\alpha-p}{p(1-p)A(\alpha,p)}$, the Stirling formula and the behavior of $D(\alpha,a)$), yields the following results.

Lemma 3.2.2 *(See [AlDi], Theorem 1.1.) If Y is beta-distributed for the parameters $(\alpha m,(1-\alpha)m)$ then*

$$\mathrm{Prob}[Y \geq p] = e^{S^m(\alpha)}\int_{-\infty}^{A(\alpha,p)}\sqrt{\frac{m}{2\pi}}e^{-\frac{m}{2}a^2}D(\alpha,a)da \tag{3.2.5}$$

where $S^m(\alpha)$ is very small for large m, namely $S^m(\alpha)\simeq(\frac{1}{12m})(1-\frac{1}{\alpha}-\frac{1}{1-\alpha})\to 0$ as $m\to\infty$.

Corollary 3.2.3 *(See [AlDi], Corollary 1.) Let Z be a standard Gaussian random variable, and let Y_m be Beta-distributed with parameters $(\alpha m,(1-\alpha)m)$. For $m\to\infty$ and for fixed p*

(i) If $p>\alpha$ then

$$\lim_{m\to\infty}\frac{\mathrm{Prob}[Y_m\geq p]}{\mathrm{Prob}[Z\leq\sqrt{m}A(\alpha,p)]} = D(\alpha,A(\alpha,p)). \tag{3.2.6}$$

(ii) If $p<\alpha$ then

$$\lim_{m\to\infty}\frac{\mathrm{Prob}[Y_m\leq p]}{\mathrm{Prob}[Z\geq\sqrt{m}A(\alpha,p)]} = D(\alpha,A(\alpha,p)). \tag{3.2.7}$$

(Note that the case $p<a$ is not covered in [AlDi], Corollary 1.)

Following [AlDi], another result can be obtained as follows (we use it in section 3.2.4).

Proposition 3.2.4 *(See [AlDi], Theorems 2.1' and 2.1".) Assume Y is a beta-distributed random variable for the parameters $(\alpha m,(1-\alpha)m)$, and denote $\alpha'=\frac{\alpha m}{m-1}$ and $\alpha''=\frac{\alpha m-1}{m-1}$. Then*

$$\begin{aligned}\mathrm{Prob}[Y &\leq p]\geq\mathrm{Prob}[Z\geq\sqrt{m-1}A(\alpha',p)],\\ \mathrm{Prob}[Y &\leq p]\leq\mathrm{Prob}[Z\geq\sqrt{m-1}A(\alpha'',p)].\end{aligned}$$

Proof of Theorem 3.2.1 We use Corollary 3.2.3 with $m = n/2$, $\alpha = (1-\lambda)$, and $p = \sin^2 \varepsilon$. Again, let Z be a standard Gaussian random variable. First examine the case $p > 1 - \lambda$. Denote by Φ the Gaussian distribution function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$. Then $\mathrm{Prob}[Z \leq \sqrt{\frac{n}{2}} A((1-\lambda), p)] = \Phi(\sqrt{\frac{n}{2}} A((1-\lambda), p))$. Notice that in this case $A((1-\lambda), p)$ is negative. Differentiation easily yields the following approximation for the Gaussian integral for y positive.

$$\frac{1}{\sqrt{2\pi}} \frac{1}{y^{-1}+y} e^{-\frac{y^2}{2}} \leq \frac{1}{\sqrt{2\pi}} \int_{y}^{\infty} e^{-\frac{1}{2}x^2} dx \leq \frac{1}{\sqrt{2\pi}} \frac{1}{y} e^{-\frac{y^2}{2}}. \tag{3.2.8}$$

This approximation implies that

$$\mathrm{Prob}[Z \leq \sqrt{\frac{n}{2}} A((1-\lambda), p)] \simeq \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{n}{2}(A^2((1-\lambda,p)/2))}}{\sqrt{n/2}(-A((1-\lambda), p))}$$

(here, and throughout this section both λ and ε ($=\arcsin\sqrt{p}$) are fixed). A direct substitution yields

$$D((1-\lambda), A((1-\lambda), p)) = \frac{\sqrt{\lambda(1-\lambda)}}{(1-\lambda)-p} A((1-\lambda), p).$$

Therefore, using Corollary 3.2.3, we get the estimate

$$\begin{aligned}
\mathrm{Prob}[Y &\geq p] \\
&\simeq \Phi\left(\sqrt{\frac{n}{2}} A((1-\lambda), p)\right) D((1-\lambda), A((1-\lambda), p)) \\
&\simeq \frac{1}{\sqrt{n\pi}} \frac{\sqrt{\lambda(1-\lambda)}}{p-(1-\lambda)} e^{-\frac{n}{2}(A^2((1-\lambda),p)/2)}.
\end{aligned}$$

In other words, if $\sin^2 \varepsilon > (1-\lambda)$ then

$$\begin{aligned}
\mathrm{Prob}[Y \leq \sin^2 \varepsilon] &= 1 - \mathrm{Prob}[Y \geq \sin^2 \varepsilon] \\
&\simeq 1 - \frac{1}{\sqrt{n\pi}} \frac{\sqrt{\lambda(1-\lambda)}}{\sin^2 \varepsilon - (1-\lambda)} e^{-\frac{n}{2}(A^2((1-\lambda),\sin^2 \varepsilon)/2)}.
\end{aligned}$$

This completes the proof in the case $\sin^2 \varepsilon > (1-\lambda)$. The case $\sin^2 \varepsilon < (1-\lambda)$ follows, in fact, from the former case. Assume $\sin^2 \varepsilon < (1-\lambda)$, and let Y' be a beta-distributed random variable with parameters $(\lambda n/2, (1-\lambda)n/2)$. Then Y' has the same distribution as $1-Y$ and therefore (since now $\cos^2 \varepsilon > \lambda$), we get

$$\begin{aligned}
\mathrm{Prob}[Y \leq \sin^2 \varepsilon] &= \mathrm{Prob}[Y' \geq \cos^2 \varepsilon] \\
&\simeq \frac{1}{\sqrt{n\pi}} \frac{\sqrt{\lambda(1-\lambda)}}{(\cos^2 \varepsilon - \lambda)} e^{-\frac{n}{2}(A^2(\lambda,\cos^2 \varepsilon)/2)}.
\end{aligned}$$

Since $A(t,s) = A(1-t, 1-s)$, the proof is complete. $\square$

3.2.4 The case of λ and ε not necessarily fixed

The estimates in the previous section apply to the case n $\to\infty$ with λ and ε fixed. In many problems, however, we would like to allow λ to change with n. To cover such instances we use Proposition 3.2.4. This proposition gives estimates from above and from below, rather than limit estimates, for the required probabilities. The expressions we get resemble the estimates attained for λ and ε fixed as follows. We first give Theorem 3.2.5, which is in the most general form. In Theorem 3.2.6 we place an extra condition which simplifies the estimates. This condition can be relaxed, to form a condition which hold in all cases relevant to us in this paper. The result is given in Theorem 3.2.7.

Theorem 3.2.5 *Let $n \geq 0$, $3 \leq k \leq n-3$, $\lambda = \frac{k}{n}$, and let E_k be a k-dimensional subspace of $\mathbb{R}^n$. For $\varepsilon > 0$, let $\mu((E_k)_\varepsilon)$ be the Haar measure on the sphere S^{n-1} of all the points within geodesic distance smaller than ε of E_k. Denote $l = \frac{\sin^2\varepsilon}{1-\lambda}$, and $l' = \frac{\cos^2\varepsilon}{\lambda}$. Then there exist positive constants $c_{\lambda,n}$ and $c'_{\lambda,n}$, bounded by an absolute constant M, such that:*

(i) If $\sin^2\varepsilon < 1-\lambda$ then

$$\frac{1}{\sqrt{2\pi}}\frac{e^{-u-c'_{\lambda,n}-\ln l'}}{\frac{1}{\sqrt{u+c'_{\lambda,n}+\ln l'}}+\sqrt{u+c'_{\lambda,n}+\ln l'}} \leq \mu((E_k)_\varepsilon) \leq \frac{1}{\sqrt{2\pi}}\frac{e^{-u-c_{\lambda,n}-\ln l}}{\sqrt{u+c_{\lambda,n}+\ln l}}. \tag{3.2.9}$$

(ii) If $\sin^2\varepsilon > 1-\lambda$ then

$$\frac{1}{\sqrt{2\pi}}\frac{e^{-u-c_{\lambda,n}-\ln l}}{\sqrt{u+c_{\lambda,n}+\ln l}} \leq 1-\mu((E_k)_\varepsilon) \leq \frac{1}{\sqrt{2\pi}}\frac{e^{-u-c'_{\lambda,n}-\ln l'}}{\frac{1}{\sqrt{u+c'_{\lambda,n}+\ln l'}}+\sqrt{u+c'_{\lambda,n}+\ln l'}} \tag{3.2.10}$$

where $u = \frac{n}{2}[(1-\lambda)\ln\frac{(1-\lambda)}{\sin^2\varepsilon}+\lambda\ln\frac{\lambda}{\cos^2\varepsilon}]$.

If we restrict our discussion to the case where the terms l and l' in Theorem 3.2.5 are bounded from below and from above, we can give a simpler formulation of the result.

Theorem 3.2.6 *Let $n \geq 0$, $3 \leq k \leq n-3$, $\lambda = \frac{k}{n}$, and let E_k be a k-dimensional subspace of $\mathbb{R}^n$. Fix $l_1 > 0$ and $l_2 < \infty$, and assume that $l_1 < \frac{\sin^2\varepsilon}{1-\lambda} < l_2$, $l_1 < \frac{\cos^2\varepsilon}{\lambda} < l_2$. Then there exist absolute constants $c_1, \dots, c_4$, depending only on l_1 and l_2, such that:*

(i) If $\sin^2\varepsilon < 1-\lambda$*, then*

$$c_1\frac{e^{-u}}{\sqrt{u}} \le \mu((E_k)_\varepsilon) \le c_2\frac{e^{-u}}{\sqrt{u}}. \tag{3.2.11}$$

(ii) If $\sin^2\varepsilon > 1-\lambda$*, then*

$$1-c_3\frac{e^{-u}}{\sqrt{u}} \le \mu((E_k)_\varepsilon) \le 1-c_4\frac{e^{-u}}{\sqrt{u}} \tag{3.2.12}$$

where $u = \frac{n}{2}[(1-\lambda)\ln\frac{(1-\lambda)}{\sin^2\varepsilon} + \lambda\ln\frac{\lambda}{\cos^2\varepsilon}]$.

The restrictions of boundedness imposed on l and l' in the statement of Theorem 3.2.6 can be relaxed. In the next theorem we assume, instead, only that l and l' are between $\frac{1}{n}$ and n.

Theorem 3.2.7 *Let* $n \ge 0$, $3 \le k \le n-3$, $\lambda = \frac{k}{n}$*, and let* E_k *be a* k*-dimensional subspace of* $\mathbb{R}^n$*. Assume that* $\frac{1}{n} < \frac{\sin^2\varepsilon}{1-\lambda} < n$, $\frac{1}{n} < \frac{\cos^2\varepsilon}{\lambda} < n$*. Then there exist absolute constants* $c_1,\ldots,c_4$*, and a sequence* $\alpha_n \to 1$*, such that:*

(i) If $\sin^2\varepsilon < 1-\lambda$*, then*

$$c_1\frac{e^{-\alpha_n u}}{\sqrt{u}} \le \mu((E_k)_\varepsilon) \le c_2\frac{e^{-\alpha_n u}}{\sqrt{u}}. \tag{3.2.13}$$

(ii) If $\sin^2\varepsilon > 1-\lambda$*, then*

$$1-c_3\frac{e^{-\alpha_n u}}{\sqrt{u}} \le \mu((E_k)_\varepsilon) \le 1-c_4\frac{e^{-\alpha_n u}}{\sqrt{u}} \tag{3.2.14}$$

where $u = \frac{n}{2}[(1-\lambda)\ln\frac{(1-\lambda)}{\sin^2\varepsilon} + \lambda\ln\frac{\lambda}{\cos^2\varepsilon}]$.

Proof of Theorems 3.2.5, 3.2.6 and 3.2.7 Proposition 3.2.4 together with the approximation (3.2.8) give, for $\sin^2\varepsilon < 1-\lambda$, that

$$\frac{1}{\sqrt{2\pi}}\frac{e^{-\xi'}}{\frac{1}{\sqrt{\xi'}}+\sqrt{\xi'}} \le \mu((E_k)_\varepsilon) \le \frac{1}{\sqrt{2\pi}}\frac{e^{-\xi}}{\sqrt{\xi}},$$

and for $\sin^2\varepsilon > 1-\lambda$, that

$$1-\frac{1}{\sqrt{2\pi}}\frac{e^{-\xi'}}{\sqrt{\xi'}} \le \mu((E_k))_\varepsilon \le 1-\frac{1}{\sqrt{2\pi}}\frac{e^{-\xi}}{\frac{1}{\sqrt{\xi}}+\sqrt{\xi}},$$

where

$$\xi = \frac{n}{2}\left((1-\lambda)(1-\frac{2}{(1-\lambda)n})\ln(\frac{(1-\lambda)(1-\frac{2}{(1-\lambda)n})(1+\frac{2}{(n-2)})}{\sin^2\varepsilon}) + \lambda\ln(\frac{\lambda(1+\frac{2}{n-2})}{\cos^2\varepsilon})\right),$$

$$\xi' = \frac{n}{2}\left((1-\lambda)\ln(\frac{(1-\lambda)(1+\frac{2}{n-2})}{\sin^2\varepsilon}) + \lambda(1-\frac{2}{\lambda n})\ln(\frac{\lambda(1-\frac{2}{\lambda n})(1+\frac{2}{n-2})}{\cos^2\varepsilon})\right).$$

Rearrangement of terms yields

$$\xi = u + \ln(1+\frac{2}{n-2})^{\frac{n-2}{2}} + \ln(1-\frac{2}{(1-\lambda)n})^{\frac{(1-\lambda)n}{2}} + \ln\frac{\sin^2\varepsilon}{(1-\lambda-\frac{2}{n})},$$

$$\xi' = u + \ln(1+\frac{2}{n-2})^{\frac{n-2}{2}} + \ln(1-\frac{2}{\lambda n})^{\frac{\lambda n}{2}} + \ln\frac{\cos^2\varepsilon}{\lambda-\frac{2}{n}}$$

when u is spelled out in the statement of the Theorems. Restriction to the case $(1-\lambda)n \geq 3$ and $\lambda n \geq 3$, implies the existence of positive constants $c_{\lambda,n}$ and $c'_{\lambda,n}$, bounded by a constant M such that

$$\xi = u + c_{\lambda,n} + \ln l,$$

$$\xi' = u + c'_{\lambda,n} + \ln l'.$$

This proves Theorem 3.2.5. In the case of Theorem 3.2.6, for large enough n, ξ and ξ' are bigger than 1 (actually they are tending to infinity) and hence the additional $\frac{1}{\sqrt{\xi'}}$ or $\frac{1}{\sqrt{\xi}}$ in the denominator of the displayed inequalities can be ignored, resulting perhaps in a slightly larger constant. Since l and l' are assumed bounded from below and from above, we can substitute $\ln l$ and $\ln l'$ by constants as well, and the proof of Theorem 3.2.6 is also complete. To prove Theorem 3.2.7, we use Theorem 3.2.5 and the boundedness assumption in a similar way. □

3.2.5 Some special cases

In this section we provide useful expressions for the general estimates in several particular cases which are handy in applications. The computations and substitutions are straightforward, and are included for completeness.

Case 3.2.8 $\sin^2\varepsilon = (1-\lambda) \pm \lambda\delta$, *where* $\lambda \to 0$ *and* $0 < \delta < 1$ *is fixed.*

(a) If $\sin^2\varepsilon = (1-\lambda) + \lambda\delta$, *then*

$$\mu((E_k)_\varepsilon) \simeq 1 - c\frac{1}{\sqrt{\frac{n}{2}(\lambda(\ln(\frac{1}{1-\delta})-\delta)+O(\lambda^2))}}e^{-\frac{n}{2}(\lambda(\ln(\frac{1}{1-\delta})-\delta)+O(\lambda^2))}, \tag{3.2.15}$$

(b) If $\sin^2\varepsilon = (1-\lambda) - \lambda\delta$, *then*

$$\mu((E_k)_\varepsilon) \simeq \frac{c'}{\sqrt{\frac{n}{2}(\lambda(\ln(\frac{1}{1+\delta})+\delta)+O(\lambda^2))}} e^{-\frac{n}{2}(\lambda(\ln(\frac{1}{1+\delta})+\delta)+O(\lambda^2))} \tag{3.2.16}$$

where, in general c and c' depend on δ. The constants c and c' become absolute if we further assume δ to be bounded away from 1 (say if we assume $\delta < 9/10$).

In order to verify (3.2.15) and (3.2.16) we establish the boundedness condition required from the expressions $\frac{\sin^2\varepsilon}{1-\lambda}$ and $\frac{\cos^2\varepsilon}{\lambda}$ in the statement of Theorem 3.2.6. Such a bound exists for a fixed δ, and it is uniform with respect to δ bounded away from 1. Now, Theorem 3.2.6 implies the following. When $\sin^2\varepsilon = (1-\lambda)+\lambda\delta$,

$$\mu((E_k)_\varepsilon) \quad \simeq \quad 1 - \frac{c}{\sqrt{-\frac{n}{2}(A^2((1-\lambda),\sin^2\varepsilon)/2)}} e^{-\frac{n}{2}(A^2((1-\lambda),\sin^2\varepsilon)/2)}.$$

We can make use of the equality

$$\begin{aligned}\frac{1}{2}A^2((1-\lambda),(1-\lambda)+\lambda\delta) &= (1-\lambda)\ln(\frac{1-\lambda}{1-\lambda+\lambda\delta}) + \lambda\ln(\frac{1}{1-\delta}) = \\ &= \lambda(\ln(\frac{1}{1-\delta})-\delta)+O(\lambda^2).\end{aligned}$$

and get the estimate, thus verify, (3.2.15).

When $\sin^2\varepsilon = (1-\lambda)-\lambda\delta$ (namely $\cos^2\varepsilon = \lambda(1+\delta)$), we look at

$$\mu((E_k)_\varepsilon) \quad = \quad \mathrm{Prob}[Y' > \cos^2\varepsilon] \simeq \frac{c'}{\sqrt{\frac{n}{2}(A^2(\lambda,\cos^2\varepsilon)/2)}} e^{-\frac{n}{2}(A^2(\lambda,\cos^2\varepsilon)/2)}.$$

Now we make use of the equality

$$\begin{aligned}\frac{1}{2}A^2(\lambda,\lambda(1+\delta)) &= \lambda\ln(\frac{1}{1+\delta}) + (1-\lambda)\ln(\frac{1-\lambda}{1-\lambda-\lambda\delta}) = \\ &= \lambda(\delta+\ln(\frac{1}{1+\delta}))+O(\lambda^2)\end{aligned}$$

and deduce estimate (3.2.16).

Remark If in the previous example we assume that δ is small, the terms in (3.2.15) and (3.2.16) are of order $e^{-n\delta^2}$. In the classical concentration result a term $e^{-n\varepsilon^2}$ appears. We note that although there is a square in both expressions, ε^2 and δ^2, they originate differently. In classical concentration the square appears when one computes the modulus of convexity of the euclidean ball. In our setting the square is a result of the behavior of the term u in Theorem 3.2.5 for small δ. This is the reason that in Case 3.2.11 below, when we use the formula in Theorem 3.2.5 for large δ, we get completely different results. Indeed, the expression $(\ln(\frac{1}{1-\delta}) - \delta)$, which for small δ behaves like δ^2, explodes when $\delta \to 1$.

Case 3.2.9 $\sin^2\varepsilon = (1-\lambda) \pm (1-\lambda)\delta$, *where* $\lambda \to 1$ *and* $0 < \delta < 1$ *is fixed.*

(a) If $\sin^2\varepsilon = (1-\lambda) + (1-\lambda)\delta$, *then*

$$\mu((E_k)_\varepsilon) \simeq 1 - \frac{ce^{-\frac{n}{2}((1-\lambda)(\ln(\frac{1}{1+\delta})+\delta)+O((1-\lambda)^2))}}{\sqrt{\frac{n}{2}((1-\lambda)(\ln(\frac{1}{1+\delta})+\delta)+O((1-\lambda)^2))}}, \tag{3.2.17}$$

(b) If $\sin^2\varepsilon = (1-\lambda) - (1-\lambda)\delta$, *then*

$$\mu((E_k)_\varepsilon) \simeq \frac{c'e^{-\frac{n}{2}((1-\lambda)(\ln(\frac{1}{1-\delta})-\delta)+O((1-\lambda)^2))}}{\sqrt{\frac{n}{2}((1-\lambda)(\ln(\frac{1}{1-\delta})-\delta)+O((1-\lambda)^2))}} \tag{3.2.18}$$

where, again, in general c *and* c' *depend on* δ, *but become absolute if* δ *is assumed to be bounded away from* 1.

Verifying (3.2.17) and (3.2.18) follows the same lines as Case 3.2.8.

Case 3.2.10 $\sin^2\varepsilon = (1-\lambda) \pm \delta$, *where* $0 < \lambda < 1$ *is fixed and* $\delta \to 0$.

(a) If $\sin^2\varepsilon = (1-\lambda) + \delta$, *then*

$$\mu((E_k)_\varepsilon) \simeq 1 - \frac{c}{\sqrt{\frac{n}{2}(\frac{\delta^2}{2\lambda(1-\lambda)} + O(\delta^3))}} e^{-\frac{n}{2}(\frac{\delta^2}{2\lambda(1-\lambda)}+O(\delta^3))}, \tag{3.2.19}$$

(b) If $\sin^2\varepsilon = (1-\lambda) - \delta$, *then*

$$\mu((E_k)_\varepsilon) \simeq \frac{c'}{\sqrt{\frac{n}{2}(\frac{\delta^2}{2\lambda(1-\lambda)} + O(\delta^3))}} e^{-\frac{n}{2}(\frac{\delta^2}{2\lambda(1-\lambda)}+O(\delta^3))} \tag{3.2.20}$$

where in general the constants c *and* c' *depend on* λ. *If* λ *is assumed to be bounded away both from* 0 *and from* 1, *the constants become absolute.*

In order to verify (3.2.19) and (3.2.20) we establish again the boundedness conditions for the expressions $\frac{\sin^2\varepsilon}{1-\lambda}$ and $\frac{\cos^2\varepsilon}{\lambda}$ in the statement of Theorem 3.2.6. Such a bound exists for a fixed λ, and it is uniform with respect to λ bounded away from 0 and from 1. Now, Theorem 3.2.6 implies the following.

When $\sin^2\varepsilon = (1-\lambda)+\delta$,

$$\mu((E_k)_\varepsilon) \simeq 1 - \frac{c}{\sqrt{\frac{n}{2}(A^2((1-\lambda),\sin^2\varepsilon)/2)}} e^{-\frac{n}{2}(A^2((1-\lambda),\sin^2\varepsilon)/2)}.$$

We can make use of the equality

$$\begin{aligned}\frac{1}{2}A^2((1-\lambda),(1-\lambda)+\delta) &= (1-\lambda)\ln(\frac{1-\lambda}{1-\lambda+\delta}) + \lambda\ln(\frac{\lambda}{\lambda-\delta}) = \\ & \frac{1}{2}\frac{\delta^2}{\lambda(1-\lambda)} + O(\delta^3).\end{aligned}$$

and get the desired estimate (3.2.19). The second case, (3.2.20), is verified in the same way.

Case 3.2.11 $\sin^2\varepsilon = (1-\lambda)-(1-\lambda)\delta$, *where* λ *is close to* 1. *(Note that the assumption here is the same as in Case 3.2.9, except that here no conditions are imposed on* δ. *We make use of this specific condition in the proof of Remark 3.2.13 below.) In this case we get an inequality as follows.*

$$\mu((E_k)_\varepsilon) \le c\frac{e^{-(1-\lambda)(\ln(\frac{1}{1-\delta})-\delta)n/2+\ln(\frac{1}{1-\delta})}}{\sqrt{(1-\lambda)(\ln(\frac{1}{1-\delta})-\delta)n/2+\ln(\frac{1}{1-\delta})}}. \tag{3.2.21}$$

In order to verify (3.2.21) we use Theorem 3.2.5. (We cannot use, as we did in the previous cases, Theorem 3.2.6, since the boundedness conditions in Theorem 3.2.6 do not hold uniformly for δ close to 1.) The terms u, l and l' in Theorem 3.2.5 are in this case

$$\begin{aligned} u &= \frac{n}{2}((1-\lambda)(\ln(\frac{1}{1-\delta})-\delta) + \frac{1}{2}(1-\lambda)^2\delta^2/\lambda^2 - \frac{1}{3}(1-\lambda)^3\delta^3/\lambda^3 + \ldots) \\ l &= \frac{\sin^2\varepsilon}{1-\lambda} = 1-\delta, \\ l' &= \frac{\cos^2\varepsilon}{\lambda} = 1 + \frac{1-\lambda}{\lambda}\delta. \end{aligned}$$

Therefore

$$u + \ln l \ge \frac{n}{2}(1-\lambda)(\ln(\frac{1}{1-\delta})-\delta) + \ln(1-\delta).$$

Inserting this inequality in Theorem 3.2.5 implies (3.2.21).

Similar inequalities can be deduced for the case $\sin^2\varepsilon = (1-\lambda)+(1-\lambda)\delta$, and likewise for the other inequality in Theorem 3.2.5. We leave out the details.

3.2.6 An application to a projection problem

Consider the orthogonal projection of a random (uniformly distributed) point x in the $(n-1)$-dimensional sphere onto a k-dimensional subspace. The expectation of $|Px|$, the euclidean norm of the projection Px, is close to $\sqrt{\frac{k}{n}}$. The reason is that the expectation of the square of the euclidean norm is exactly $\frac{k}{n}$. Moreover, the set of points x for which $|Px|$ is close to $\sqrt{\frac{k}{n}}$ has a large measure. This gives rise to the following general question, several aspects of which we address in this section.

Question 3.2.12 *Given a set of cardinality N on S^{n-1}, does there exist a subspace of dimension k such that the euclidean norms of the elements in the projection of the set on the subspace are all not far from $\sqrt{\frac{k}{n}}$?*

An answer to this question can be one of two types. First, for a given N (which depends on the dimension n), one may provide an estimate $\Delta(N)$, such that whenever a set of cardinality N on the sphere is given, a k-dimensional subspace exists such that all the projections on the subspace of the points in the set have euclidean norms which are $\Delta(N)$ close to $\sqrt{\frac{k}{n}}$. Second, for a given degree of closeness Δ, one may estimate the maximal cardinality $N(\Delta)$ such that whenever a set of cardinality $N(\Delta)$ is given, a k-dimensional subspace exists such that all the projections on the subspace of the points in the set have euclidean norms which are Δ-close to $\sqrt{\frac{k}{n}}$. In both cases we are interested in the behavior of the estimates for large dimension n. We refer to an answer of the first type as an isomorphic answer and to the second type, for a fixed Δ independent of dimension, as a Δ-isometric answer. In the first part of this section we give a general, however not transparent, answer to Question 3.2.12, addressing both types of estimates. In the second part of the section we provide a more concrete form of the answer in a particular case (which we then use in the next section).

Our way to establish the existence of a subspace with a certain property, is to show that the measure of subspaces in $G_{n,k}$ with this property is positive. ($G_{n,k}$ is, as customary, the Grassman manifold of k-dimensional subspaces of $\mathbb{R}^n$, endowed with the normalized Haar measure.)

To provide our answer to Question 3.2.12, we need the following observations. First, the measure of subspaces in $G_{n,k}$ with a certain property is the same as the measure of orthogonal transformations $U \in O(n)$ such that $U(E_0)$ has this property, where E_0 is some fixed subspace of dimension k. (Here $O(n)$ is the group of orthogonal transformations on $\mathbb{R}^n$ endowed with the normalized

Haar measure.) Second, consider a set of points $\{x_i\}_{i=1}^N$ in $\mathbb{R}^n$. Placing a restriction on the norms of their projections on $U(E_0)$, is equivalent to placing the same restriction on the norms of the projections to E_0 of the set $\{U^{-1}x_i\}_{i=1}^N$. This is implied by the equalities $P_{UE_0}x = P_{UE_0}UU^{-1}x = UP_{E_0}U^{-1}x$. Third, consider again a set of points $\{x_i\}_{i=1}^N$ in $\mathbb{R}^n$. For a given set $I \subset [0,1]$, if $1 - \sum_{i=1}^N \mu\{U \in O(n) : P_{E_0}U^{-1}x_i \notin I\} > 0$, then the measure of the family of orthogonal transformations satisfying $|P_{E_0}U^{-1}x_i| \in I$ for every $1 \le i \le N$, is positive.

For $k = (1-\lambda)n$ and for $\Delta > 0$ define

$$\mu_0 = \mu\{x \in S^{n-1} : ||P_{E_k}x| - \sqrt{\frac{k}{n}}| < \Delta\}. \tag{3.2.22}$$

Taking into account these three observations, we see that if $N < 1/\mu_0$, then for every set $\{x_i\}_{i=1}^N$ in S^{n-1} there exists a subspace E_k of dimension k such that for every x_i, $i = 1 \ldots n$, $|P_{E_k}x_i - \sqrt{\frac{k}{n}}| < \Delta$. The above enables us to establish relations $N(\Delta)$ and $\Delta(N)$ as an answer to Question 3.2.12. Indeed, given Δ, we can use sections 3.2.3, 3.2.4 and 3.2.5 to estimate $\mu_0 = \mu_0(\Delta)$, using the following transparent equality

$$\begin{aligned}\mu_0 &= \mu\{x : \sqrt{1-\lambda} - \Delta < |P_{E_k}x| < \sqrt{1-\lambda} + \Delta)\} \\ &= \mu\{x : \mathrm{d}(x, E_{\lambda n})^2 < (\sqrt{1-\lambda} + \Delta)^2\} - \mu\{x : \mathrm{d}(x, E_{\lambda n})^2 < (\sqrt{1-\lambda} - \Delta)^2\},\end{aligned}$$

and then have the estimate $N(\Delta) = 1/\mu_0(\Delta)$. To get an estimate for $\Delta(N)$, we do the reverse. In Proposition 3.2.14 we use the above scheme for a special choice of N.

Remark 3.2.13 *The preceding derivations, and in particular the estimate for $N(\Delta)$, are close in spirit to the Johnson-Lindenstrauss Lemma, see [JoLi]. The Lemma gives an estimate for the smallest dimension $k(n)$ such that any subset of cardinality n of l_2^n can be $(1+\epsilon)$-isometrically embedded into $l_2^{k(n)}$. Such an embedding can be realized by projecting into a well chosen subspace of the appropriate dimension, and by dividing then the images by $\sqrt{\lambda}$. The reason the method works is the following. Instead of the set of points, say $\mathcal{N}$, in l_2^n, one considers the set $\mathcal{N}' = \{\frac{x_i - x_j}{|x_i - x_j|}\}$ of normalized differences. To this end a restriction has to be imposed on the cardinality of $\mathcal{N}$, or equivalently on the dimension k (the exact restriction can easily be computed). Then the existence of a subspace E_k such that for every $y \in \mathcal{N}'$ the euclidean norm of it's projection onto E_k is close to $\sqrt{\lambda}$ is guaranteed. This insures that the distance between Px_i and Px_j is close to $\sqrt{\lambda}$ times the distance between x_i and x_j. In other words, the* *relative distances do not change much. Thus, the resemblance to the derivations in the present paper is apparent. However, while in the Johnson-Lindenstrauss Lemma the embedding was required to be an ε-isometry, the present paper offers a general isomorphic version.*

Remark The following estimate was used in an article by Milman and Pajor [MPa].

$$\mathrm{Prob}[T \in O(n) : |P_{E_k}Tx| > \xi|x|] \leq \left(e(1-\xi^2)\frac{1}{1-\lambda}\right)^{(1-\lambda)n/2}. \tag{3.2.23}$$

In the terminology of the present paper the left hand side of (3.2.23) has the following form.

$$\mathrm{Prob}[x \in S^{n-1} : \mathrm{dist}^2_{\|\cdot\|_2}(x, E_k) < 1 - \xi^2].$$

Estimates for this expression better than (3.2.23) can be obtained from the results in Sections 3.2.3-3.2.5. Such estimates would be asymptotically accurate. Moreover, (3.2.23) has meaning only when $\xi^2 > \lambda(1 + \frac{1-\lambda}{\lambda}(1-\frac{1}{e}))$, while using the technique of the present paper yields an answer whenever $\xi^2 > \lambda$.

We turn now to the special case for which we get a more concrete answer to Question 3.2.12. Suppose that e^{ck} points in S^{n-1} are projected into a k-dimensional subspace E_k. If c is large, it could be that for no subspace E_k all the norms of the projections are close to $\sqrt{\frac{k}{n}}$. We can, however, choose E_k such that all of the projections do not enter some small neighborhood of 0, namely, we can establish some kind of isomorphic result. The precise result is as follows.

Proposition 3.2.14 *For any number c, there exist an $\varepsilon(c)$ (for instance $\varepsilon(c) = e^{-(2c+1)}$ will work) and $C(c)$, such that the following holds for every $\lambda = \frac{k}{n}$ fixed. For n large enough, whenever a set $\mathcal{N}$ in S^{n-1} is of cardinality $|\mathcal{N}| = e^{ck}$, there exists a subspace E of dimension k such that*

$$\sqrt{\lambda}\varepsilon(c) < |P_E x| < \sqrt{\lambda}C(c) \tag{3.2.24}$$

for every $x \in \mathcal{N}$. Moreover, by taking a small enough $\varepsilon(c)$ one can ensure that the measure of the set of subspaces in $G_{n,k}$ satisfying, say, the left hand side of (6.3), is arbitrarily close to 1 (for instance, for $\varepsilon(c) = \frac{1}{2}e^{-(2c+1)}$ the latter measure is larger than $1 - e^{\frac{1}{2}(1-\lambda)n}$).

Proof We will show how to establish the left hand side inequality, with $\varepsilon(c) = e^{-(c+1)}$. This is the only part we use in the application. The right hand side is attained in a similar way, and for the two inequalities to hold together we reduce $\varepsilon(c)$ to the magnitude mentioned in the statement of the theorem and increase $C(c)$ accordingly; thus we may also have both sides with a large measure, of course. To prove the left hand side, first note that the projections of N points do not enter an ε-neighborhood of 0 if and only if they all stay within distance more than ε from $E_k^\perp$. Using the same reasoning as in the general answer above, we find that for the latter condition to hold it is enough that N be smaller than $1/\mu((E_{n-k})_\varepsilon)$. For $1-\lambda = \frac{k}{n}$ small, and when $\sin^2\varepsilon = (1-\lambda)(1-\delta)$ we use Case 3.2.11 to get that the inequality

$$N \le \frac{\sqrt{\frac{k}{2}(\ln(\frac{1}{1-\delta}) - \delta) - \ln(\frac{1}{1-\delta})}}{c'} e^{\frac{k}{2}(\ln(\frac{1}{1-\delta})-\delta)-\ln(\frac{1}{1-\delta})} \tag{3.2.25}$$

ensures that $N < 1/\mu((E_{n-k})_\varepsilon)$. Since the function $(\ln(\frac{1}{1-\delta}) - \delta)$ can be arbitrarily large for δ close to 1, we see that for any number N of points on the sphere, there exists a $0<\delta<1$, and a subspace E_k of dimension k such that the projections, on E_k, of all the points, are out of a $\sqrt{(1-\lambda)(1-\delta)}$-neighborhood of 0. For the specific case $N = e^{ck}$ mentioned in the statement of the Proposition, we identify $\delta = \delta(c)$ for which (3.2.25) holds. To this end, for k large enough, it is enough to ask that

$$ck \le \frac{k}{2}(\ln(\frac{1}{1-\delta}) - \delta) - \ln(\frac{1}{1-\delta}),$$

or equivalently that $1-\delta \sim e^{-2c}$. This identifies the desired δ. In particular the size of the neighborhood of 0 is

$$\varepsilon(c) = \sqrt{(1-\lambda)(1-\delta)} \sim \sqrt{(1-\lambda)}e^{-c}.$$

This verifies the first claim of Proposition 3.2.14. By choosing a smaller $\varepsilon(c)$, and following the same line of proof, one can verify that the inequality (3.2.24) is satisfied for a large measure of subspaces, and complete the proof.

We may apply Proposition 3.2.14 to the set of normalized differences $\mathcal{N}' = \{\frac{x-y}{|x-y|} : x, y \in \mathcal{N}\}$, and get the following.

Corollary 3.2.15 *For any number c, there exist $\varepsilon'(c)$ (for instance $\varepsilon'(c) = e^{-(2c+1)}$) and $C'(c)$ such that the following holds for every $\lambda = \frac{k}{n}$ fixed. For n large enough, whenever a set $\mathcal{N}$ in R^n of cardinality $|\mathcal{N}| = e^{ck}$, there exists a subspace E of dimension k such that*

$$\sqrt{\lambda}\varepsilon'(c)|x-y| < |P_E x - P_E y| < \sqrt{\lambda}C'(c)|x-y| \tag{3.2.26}$$

for every $x, y \in \mathcal{N}$.

3.2.7 An application to the duality of entropy numbers

In this section we use Corollary 3.2.15 to answer a question regarding entropy numbers. We remark that in Chapter 1 of this thesis covering numbers were introduced, and the duality of entropy numbers conjecture was proved, in the case where one of the two bodies is the euclidean ball, (which we denote by D_n, or by D when the dimension of the space is understood from the context). The result described below is a special case of the results presented in Chapter 1; however, the method below is much more elementary than those of Chapter 1. Moreover, historically this work's results preceded the results of Chapter 1, and were part of the accumulation of facts which led us to our general solution, described in Chapter 1 (though by now, with the simplified proofs, this step is no longer needed).

Recall the covering number, $N(K,T)$, of K by T, where K and T are two convex bodies in R^n, given by

$$N(K,T) = \min\{N : \exists \{x_i\}_{i=1}^N \subset \mathbb{R}^n, K \subset \cup_{i=1}^N \{x_i\} + T\}.$$

As explained in Chapter 1, the general duality conjecture suggests the existence of absolute constants α and β such that for every pair of centrally symmetric convex bodies T and K the inequality

$$N(K,T) \leq (N(T^\circ, \alpha K^\circ))^\beta \tag{3.2.27}$$

holds, where T° and K° denote the polar bodies of T and K respectively.

In Theorem 3.2.16 below we verify the duality conjecture in the special case where $T = D_n$ and $\log N(K,D) = \gamma n$, where n is the dimension of the space. We need the result from [KöM], which was given in Chapter 1 as Lemma 1.4.6, and states that there exists a universal constant C (there it was denoted C') such that for any n, and any two convex bodies $K, T \subset R^n$

$$\frac{1}{C} \leq \left(\frac{N(K,T)}{N(T^\circ, K^\circ)}\right)^{\frac{1}{n}} \leq C. \tag{3.2.28}$$

We prove here the following

Theorem 3.2.16 *Assume that* $N(K,D) = 2^{c_1 n}$.

(a) If $c_1 > 2\log_2 C$, *then*

$$N(D, K^\circ) \geq 2^{c_1 n/2}, \tag{3.2.29}$$

(b) If $c_1 < 2\log_2 C$, *then*

$$N(D, (\frac{1}{4\sqrt{\log_2 C}})\sqrt{c_1} 2^{-c_1} K^\circ) \geq 2^{c_1 n/2}, \tag{3.2.30}$$

where C is the absolute constant guaranteed from (3.2.28).

Proof of Theorem 3.2.16 Case (a) follows immediately from (3.2.28). Indeed, since we assume that $N(K,D) = 2^{c_1 n}$, it follows from equation (3.2.28) that $N(D,K^\circ) \geq (2^{c_1}/C)^n$. This together with $c_1 > 2\log_2 C$ implies (3.2.29). It is not possible to use the same argument in case (b) since in this case c_1 is too small. Our strategy to handle case (b) is to use Corollary 3.2.15 in order to reduce the dimension, and only then use (3.2.28). The original dimension can then be recovered with the help of purely geometric arguments. Here are the details. Take a 1-separated set $\{x_1,\ldots,x_{2^{c_1 n}}\}$ in K. Such a large separated set exists. For example, a maximal 1-separated set is always a 1-covering, and since assuming that $N(K,D) = 2^{c_1 n}$, it follows that a 1-covering collection has at least $2^{c_1 n}$ elements. Chose a dimension $k < n$ with $2^{c_1 n} > C^k$. By Corollary 3.2.15, for any given set of $2^{c_1 n}$ points, we can find a subspace E of dimension k, such that after projecting the set of points into E, the mutual distances do not shrink by more than $\sqrt{\frac{k}{n}}\varepsilon'(c_1)$. This means that the new set, $\{P_E x_i : 1 \leq i \leq 2^{c_1 n}\} \subset P_E K$, is ε_0-separated, for $\varepsilon_0 = \sqrt{\frac{k}{n}}e^{-2c_1+1}$. Since every $z_j + \frac{\varepsilon_0}{2}D$ is of diameter ε_0 and therefore can include only one point of this new set, we get

$$N(P_E K, \frac{\varepsilon_0}{2}(D \cap E)) \geq 2^{c_1 n} = 2^{(c_1 \frac{n}{k})k}.$$

Now, since in case (b) we have $2^{c_1 \frac{n}{k}} > C$, we can apply (3.2.28) to the latter formula and extract the following meaningful estimate.

$$N((D\cap E)^\circ, \frac{\varepsilon_0}{2}(P_E K)^\circ) \geq 2^{(c_1 \frac{n}{k} - \log_2 C)k}$$

Using the facts that $(P_E K)^\circ = K^\circ \cap E$, and that $(D\cap E)^\circ = D \cap E$, and the property that for every T symmetric and convex $N(D\cap E, T\cap E) \leq N(D, \frac{1}{2}T)$, we get

$$N(D, \frac{\varepsilon_0}{4}K^\circ) \geq 2^{(c_1 \frac{n}{k} - \log_2 C)k}.$$

Applying $k = \frac{c_1 n}{2\log_2 C}$ to the latter formula verifies (3.2.30). This completes the proof. □

We remark on some extensions and comparisons:

Remark The reasoning in the preceding proof can be applied without an assumption on the cardinality of $N(K,D)$. The estimate that would emerge is

$$N(D, c\sqrt{\frac{\ln N(K,D)}{n}}K^\circ) \geq (N(K,D))^2. \tag{3.2.31}$$

In the language of entropy numbers inequality (3.2.31) has a more appealing form (recall the definition of entropy numbers for $u : X \to Y$, namely, $e_k(u) = \inf\{\varepsilon : N(B(Y), \varepsilon B(X)) \le 2^k\}$), as follows. For $u : {l_2}^n \to X$,

$$e_k(u^*) \ge c\sqrt{\frac{k}{n}} e_{2k}(u). \tag{3.2.32}$$

Note that the duality conjecture (proved in Chapter 1) means that the term $\sqrt{\frac{k}{n}}$ can be eliminated. It should be mentioned that results similar to (3.2.32) regarding the duality conjecture in the complimentary case (where the roles of K and D are interchanged) were established by V.D. Milman and S.J. Szarek in [MSz1].

Remark Theorem 3.2.16 can be compared with a result of G. Pisier, written originally for operators of rank $\le n$ (see [Pis2] Corollary 2.4). In the language of covering numbers it reads as follows. Under the condition $N(K, D) = 2^{c_1 n}$ as in Theorem 3.2.16,

$$N(D, c' \frac{c_1^2}{(1 + \log(2/c_1))} K^\circ) \ge 2^{c_1 n/2},$$

where c' is an absolute constant. Note that the inequalities (3.2.29) and (3.2.30) provide better estimates.

3.2.8 A comparison with a result of Diaconis and Freedman.

Diaconis and Freedman, [DiFr], compared the limit behavior of the following two distributions. One is the distribution of the first k coordinates of a uniformly distributed random point $(x_1, \dots, x_n)$ on the $(n-1)$-dimensional sphere of radius $\sqrt{n}$. The other is the distribution of k independent standard Gaussian random variables in $\mathbb{R}^k$. Denote the two distributions by $Q_{n,\sqrt{n},k}$ and P_1^k respectively. Diaconis and Freedman found that when $n \to \infty$ and $k = o(n)$, the variational distance between the two distributions tends to zero. A precise estimate for the convergence is

$$\left\|P_1^k - Q_{n,\sqrt{n},k}\right\| \le \frac{2(k+3)}{n-k-3}.$$

where the variational distance is defined by $||P - Q|| = 2 \sup_A |P(A) - Q(A)|$, where A runs on all measurable subsets of the probability space.

In this section we use our previous results to establish similarity of the two distributions in a different sense, and in the case $k = \lambda n$, when $0 < \lambda < 1$ is a small but fixed number. We examine a different similarity notion since we are interested in the limit behavior of the tails of the distributions. Note that both distributions concentrate strongly, as $n \to \infty$, around their expectations, with tail distributions exponentially small. Therefore, similarity of the tails would not be captured by establishing that the variational distance is small. We compare the first order terms in the exponents describing the tails. Here are the details.

The symmetry of the k coordinates in both distributions, reduces the question to a comparison between the following two one-dimensional distributions (as functions of t').

$$\begin{aligned} P'_1(t') &= \text{Prob}\,[\sum_{i=1}^{k} y_i^2 \le t' : y = (y_1, \dots, y_n) \text{ uniform on } \sqrt{n}S^{n-1}] \\ P'_2(t') &= \text{Prob}\,[\sum_{i=1}^{k} g_i^2 \le t' : g_i \text{ i.i.d. gaussians, } 1 \le i \le k] \end{aligned}$$

Denote $t = t'/k$, $x = y/\sqrt{n}$ and $P_i(t) = P'_i(tk)$. The displayed distributions get the form

$$\begin{aligned} P_1(t) &= \text{Prob}\,[\sum_{i=1}^{k} x_i^2 \le t : x = (x_1, \dots, x_n) \text{ uniform on } S^{n-1}] \\ &= \text{Prob}\,[\mathrm{d}^2(x, E_{n-k}) \le \lambda t], \\ P_2(t) &= \text{Prob}\,[\frac{1}{k}\sum_{i=1}^{k} g_i^2 \le t : g_i \text{ i.i.d. gaussians, } 1 \le i \le k]. \end{aligned}$$

We assume now that $\frac{k}{n}$ is small and estimate $P_1(t)$ using (3.2.17) and (3.2.18) in Section 3.2.5. We can also estimate $P_2(t)$ using classical results such as Cramér's theorem (see e.g. [Va]). The results of such estimates yield

(a) If $t = 1 + \delta$ then

$$\begin{aligned} P_1(t) &= 1 - e^{-\frac{k}{2}(\ln(\frac{1}{1+\delta})+\delta)+o(k)} \\ P_2(t) &= 1 - e^{-\frac{k}{2}(\ln(\frac{1}{1+\delta})+\delta)+o(k)} \end{aligned}$$

(b) If $t = 1 - \delta$ then

$$\begin{aligned} P_1(t) &= e^{-\frac{k}{2}(\ln(\frac{1}{1-\delta})-\delta)+o(k)} \\ P_2(t) &= e^{-\frac{k}{2}(\ln(\frac{1}{1-\delta})-\delta)+o(k)} \end{aligned}$$

We see that for each fixed $\delta > 0$, as $k \to \infty$ the two respective tails differ by at most $o(k)$ multiplying an exponentially (in k) small number.

3.2.9 A Remark on a recent result of M. Gromov

In a recent remarkable paper by M. Gromov [Gr] the following theorem, to which our results above are clearly connected, was proved.

Theorem 3.2.17 (Gromov) *Let $f : S^n \to \mathbb{R}^k$ be a continuous map, where S^n is the unit n-sphere. Then there exists a point $z \in \mathbb{R}^k$ such that the spherical n-volumes of the ε-neighbors of the level $Y_z = f^{-1}(z) \subset S^n$, denoted $Y_z + \varepsilon \subset S^n$, satisfy*

$$\mathrm{Vol}(Y_z + \varepsilon) \geq \mathrm{Vol}(S^{n-k} + \varepsilon) \tag{3.2.33}$$

for all $\varepsilon > 0$, where $S^{n-k} \subset S^n$ denotes an equitorial $(n-k)$-sphere.

A special case of this theorem, when $k = 1$, is the isoperimetric inequality, and in this case $z \in \mathbb{R}$ can be taken to be the Levy mean of f. The computation of the quantity $\mathrm{Vol}(S^{n-k} + \varepsilon)$ in equation (3.2.33) above was precisely the goal of Sections 3.2.3-3.2.5, and gives the following Corollary.

Corollary 3.2.18 *Let $f : S^n \to \mathbb{R}^{\lambda n}$ be a continuous map. Then there exists a point $z \in \mathbb{R}^{\lambda n}$ such that for every ε with $\sin^2 \varepsilon > \lambda$ one has*

$$\mathrm{Vol}(\{x \in S^n : \exists y \in S^n, \rho(x, y) < \varepsilon, f(y) = z\}) \geq 1 - \frac{1}{\sqrt{n\pi}} \frac{\sqrt{\lambda(1-\lambda)}}{\sin^2 \varepsilon - \lambda} e^{-\frac{n}{2} u(1-\lambda,\varepsilon)}, \tag{3.2.34}$$

where $u(1-\lambda, \varepsilon) = \lambda \ln \frac{\lambda}{\sin^2 \varepsilon} + (1-\lambda) \ln \frac{(1-\lambda)}{\cos^2 \varepsilon}$.

In particular, if one assumes that f is a Lipschitz function with Lipschitz constant C, this gives an upper estimate for the volume of the inverse image of an ε/C neighborhood of z, much like the usual concentration inequalities for functions into $\mathbb{R}$.

3.3 Sign-Projections

3.3.1 Introduction

In this section we study the behavior of the euclidean diameter of a convex body under a certain type of linear mappings called sign-projections. A sign-projection of level k is a linear mapping $S_k : \mathbb{R}^n \to \mathbb{R}^k$ of the form

$$S_k x = ((x, \xi_1), \ldots, (x, \xi_k)),$$

where for $i = 1, \ldots, k$ the vectors ξ_i are vertices of the n-dimensional cube of edge length $2/\sqrt{n}$, that is, $\xi_i = (\xi_{i,1}, \ldots, \xi_{i,n}) \in \{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$. For most choices of such $\xi_1, \dots, \xi_k$ the mapping S_k is indeed of rank k. In the special case when $\xi_1, \dots, \xi_k$ are orthonormal, S_k is an orthogonal projection. We equip the set of level k sign-projections with a probability measure in the following way: for every $i = 1, \dots, k$, the vector ξ_i is chosen uniformly in $\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$. Thus we have, for a given dimension n, 2^{nk} different sign-projections available, and each has probability 2^{-nk}.

Not much is known about the behavior of the euclidean diameter of convex sets under such mappings. Much more is known if, instead, the vectors $\xi_1, \dots, \xi_k$ are chosen to be independent Gaussian random vectors, see e.g. [DaSz]. The other well understood case is when the vectors $\xi_1, \dots, \xi_k$ are chosen to be a random *orthonormal system* in R^n. Let us explain what is known about the situation then: This is equivalent to studying projections onto random k-dimensional subspaces, i.e. onto elements in the Grassmann manifold $G_{n,k}$ endowed with the normalized Haar measure. In this case it follows from the classical Dvoretzky type theorem, as observed in [Mi1], [MSch], [Mi2], that for every convex body K in R^n there exists some critical dimension k^* (depending on K) such that, for all integers $k \geq k^*(K)$, the diameter of K shrinks by a factor close to $\sqrt{\frac{k}{n}}$ when K is projected onto a random k-dimensional subspace, and for $k \leq k^*$ the projection of K onto a random k-dimensional subspace is already, with high probability, an isomorphic euclidean ball of radius $\simeq \sqrt{\frac{k^*}{n}}\mathrm{diam}(K)$. (Whenever we use the notation $\simeq$ we mean equivalence up to universal constants independent of the body and of the dimension.) After defining the standard parameter

$$M^*(K) = \int_{S^{n-1}} \sup\{< x, u >: x \in K\} du,$$

which is half of the mean width of K (and will play a central role in this note), there is a simple formula for this critical dimension $k^*(K)$:

$$k^* = \left(\frac{M^*(K)}{\mathrm{diam}(K)}\right)^2 n. \tag{3.3.1}$$

Thus for dimension $\simeq k^*$ (and hence for all lower dimensions) for a very large measure of subspaces E in $G_{n,k}$, the projection of K onto E is close to a euclidean ball of radius $M^*(K)$. A precise formulation of the above can be found for example in [MSz1], Proposition 3.3.

We show that for $k \geq k^*$ the same decay of diameter occurs when considering random level k sign-projections instead of random orthogonal projections (and k^* is defined in (3.3.1)). The question regarding sign-projections of levels smaller than k^* remains open, and we provide some details concerning this question at the end of this note.

The central tool in the proof of our main theorem of this section is the beautiful Majorizing Measures theorem of M. Talagrand. For the precise formulation see Theorem 3.3.4 below; we emphasize that it is an estimate concerning covering numbers, in a very fine tuned way. It is a strengthening of Sudakov inequality, which says that for every ε one has $N(K, \varepsilon D) \le e^{cn(M^*/\varepsilon)^2}$ (actually asymptotically $c = 1$). Majorizing Measures theorem actually tells us that we can find a family of nets, at all levels of resolution, with some mutual good behavior. This is strongly used in our proof, where Sudakov's inequality would give a slightly weaker result.

3.3.2 Decay of norm for a single point

The first step in investigating how a random sign-projection affects the diameter of a convex body is to see how it affects the euclidean norm of a single point. Then one can hope to use this for, say, a *net* of points inside the convex body, or, as will be done here, for a *family* of nets. Another way to view this is that one first investigates the diameter decay for the simplest of all convex sets - a line segment.

The first lemma is an easy estimate, yet will serve as our main technical tool throughout the paper.

Lemma 3.3.1 *In the above notation, for a given point $x \in \mathbb{R}^n$ and any $t > 1$,*

$$\text{Prob}\ [|S_k x| > t\sqrt{\frac{k}{n}}|x|] \le \exp -(\frac{t^2 - 1 - 2\log t}{2}k) \tag{3.3.2}$$

(where $\text{Prob}(A)$ *denotes the probability of the set of level k sign-projections S_k satisfying A).*

Proof of Lemma 3.3.1. To generate the measure on S_k's we choose independently nk signs $\varepsilon_{i,j}$ for $i = 1, \dots, k$ and $j = 1, \dots, n$, and define the vectors $\xi_i = (\xi_{i,1}, \dots, \xi_{i,n})$ by $\xi_{i,j} = \varepsilon_{i,j}/\sqrt{n}$. The random operator is then defined by

$$S_k(x) = \sum_{i=1}^{k} (x, \xi_i) e_i$$

(where e_i denotes the standard unit vector basis in R^n).

We denote $\varepsilon_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,n})$. Thus ε_i is a random vector whose n coordinates are ± 1 with probabilities $1/2$, and

$$\begin{aligned}\text{Prob }[|S_k(x)| > t\sqrt{\frac{k}{n}}|x|] &= \text{Prob }[\sum_{i=1}^{k}(x,\xi_i)^2 > t^2\frac{k}{n}|x|^2] \\ &= \text{Prob }[\sum_{i=1}^{k}(x,\varepsilon_i)^2 > t^2k|x|^2].\end{aligned}$$

Notice that (denoting $x = (x_1, \ldots, x_n)$)

$$\mathbb{E}((x,\varepsilon_i)^2)) = \mathbb{E}(\sum_{j=1}^{n} x_j\varepsilon_{i,j})^2 = |x|^2$$

(where $\mathbb{E}(B)$ is the expectation of the random variable B).

We will make use of the following well known inequality which was pointed out to us by S. Bobkov: For $\lambda \in [0, \frac{1}{2})$ and $|y| = 1$,

$$\mathbb{E}(\exp(\lambda(y,\varepsilon_i)^2)) \leq \frac{1}{\sqrt{1-2\lambda}}. \tag{3.3.3}$$

For completeness we provide a proof of this inequality in the end of this note. It uses the fact that one can majorize this expression by the same expression for appropriate Gaussians, see e.g. [Ach], [Ta3].

Using Chebyshev's inequality, the independence of the ε_i's, and the above estimate, we get that for $\lambda \in (0, \frac{1}{2|x|^2})$,

$$\begin{aligned}\text{Prob }[\sum_{i=1}^{k}(x,\varepsilon_i)^2 > t^2k|x|^2] &= \text{Prob }[\exp(\lambda\sum_{i=1}^{k}(x,\varepsilon_i)^2) > \exp(\lambda t^2k|x|^2)] \\ &\leq \exp(-\lambda t^2k|x|^2)\mathbb{E}(\exp(\sum_{i=1}^{k}\lambda(x,\varepsilon_i)^2)) \\ &= \exp(-\lambda t^2k|x|^2)\mathbb{E}(\exp(\lambda(x,\varepsilon_1)^2))^k \\ &\leq \exp(-\lambda t^2k|x|^2)(\frac{1}{\sqrt{1-2\lambda|x|^2}})^k\end{aligned}$$

For $t > 1$ we optimize over λ, choosing $\lambda|x|^2 = \frac{1}{2} - \frac{1}{2t^2}$, and get that

$$\text{Prob }[|S_k(x)| > t\sqrt{\frac{k}{n}}|x|] \leq \exp-(\frac{t^2-1-2\log t}{2}k),$$

which completes the proof of (3.3.2).

Lemma 3.3.1 gives only an upper estimate for the norm. It shows that for a given point $x \in \mathbb{R}^n$ (regardless of its position), with large probability, the euclidean norm of a level k sign-projection of x is not much bigger than $\sqrt{\frac{k}{n}}$ times the euclidean norm of x. One can also show that actually this is the right decrease, and that with large probability the norm does not decrease more. The following Lemma is proved in [Ach]

Lemma 3.3.2 *In the above notation, for a given point $x \in \mathbb{R}^n$ and $0 < \varepsilon < 1$,*

$$\text{Prob }[|S_k x| < \sqrt{1-\varepsilon}\sqrt{\frac{k}{n}}|x|] \leq \exp -\frac{k}{2}(\varepsilon^2/2 - \varepsilon^3/3).$$

The main difference between the two lemmas is that in Lemma 3.3.1 we can force the probability of the event to be as small as we want by choosing a suitable (large enough) t. In this sense, Lemma 3.3.1 can be considered as isomorphic, while Lemma 3.3.2 is not.

3.3.3 Decay of diameter of a convex body

In this section we prove the main result of this note which is

Theorem 3.3.3 *There exist universal constants C_1, C_2, c_3 and c_4 such that for every dimension n, for every convex body $K \subset \mathbb{R}^n$, for $k \geq C_1 k^*$, the probability of level k sign-projections such that*

$$c_4\sqrt{\frac{k}{n}}\text{diam}(K) \leq \text{diam}(S_k K) \leq C_2\sqrt{\frac{k}{n}}\text{diam}(K)$$

is greater than $1 - 2^{-c_3 k}$, where k^* is defined in (3.3.1).

Remark: The left hand side is trivially true, because taking just one point for which the diameter is attained, the results of Section 3.2 tell us that the euclidean norm of the projection of this point, with high probability, is no less than $c_4\sqrt{\frac{k}{n}}\text{diam}(K)$. Thus we need only prove the right hand side.
Remark: In fact the theorem we prove is slightly stronger, for *every* choice of $C_1 > 0$ and $c_3 > 0$ there exist C_2 and c_4 such that the statement of the theorem holds true. Again, this fact for c_4 is immediate from Section 3.2.

The idea of the proof is to use the information from Lemma 3.3.1 on individual points in order to get an estimate for the whole convex body together. The standard way to do so is to use a covering of the body by small Euclidean balls, and operate on each ball individually, hoping that the probabilities are big enough so that many "good" events happen all at once. The standard estimate for covering numbers (how many euclidean balls are needed to cover the convex body K) is Sudakov inequality (see e.g. [Pis3], page 68), which, as it turns out, is not strong enough to prove Theorem 3.3.3. Instead we use the (stronger) theorem of M. Talagrand, called the Majorizing Measure Theorem. This theorem can be formulated in several different ways, of which the following is most suitable for our needs. We refer the reader the paper [Ta1] and to the forthcoming book [Ta2] for a proof, different formulations, and related topics.

Theorem 3.3.4 (Majorizing Measures) *There exists a universal constant C such that for any dimension n, for every convex body $K \subset \mathbb{R}^n$, there exist families of points $B_0 \subset \ldots \subset B_{m-1} \subset B_m \subset \ldots \subset K$ with cardinality $|B_m| \leq 2^{2^m}$ such that for every $x \in K$*

$$\sum_{m=0}^{\infty} d(x, B_m)\sqrt{2^m} \leq CM^*(K)\sqrt{n},$$

where $d(x, B_m)$ denotes the distance of the point x to the m^{th} family, i.e., $d(x, B_m) = \inf\{d(x, y) : y \in B_m\}$.

Proof of Theorem 3.3.3 Fix a constant $C_1 > 0$ and let $k > C_1 k^*$. We assume without loss of generality that k is a power of 2, and denote by m_0 the number such that $2^{m_0} = k$. For $x \in K$ write

$$x = v_{m_0}(x) + \sum_{m=m_0+1}^{\infty} v_m(x) - v_{m-1}(x)$$

where $v_m(x) \in B_m$ is the member of B_m closest to x. Thus, as $B_m \subset B_{m-1}$, we see that denoting $u_m(x) = v_m(x) - v_{m-1}(x)$ assures $|u_m(x)| \leq 2d(x, B_{m-1})$. We now denote $A_m = \{u_m(x) : x \in K\}$, and have $|A_m| \leq 2^{2^{m+1}}$. Rewrite

$$\begin{aligned} x &= v_{m_0} + \sum_{m=m_0+1}^{\infty} u_m(x) \\ &= v_{m_0} + \sum_{m=m_0+1}^{\infty} \left(\frac{d(x, B_{m-1})\sqrt{2^{m-1}}}{CM^*(K)\sqrt{n}} \frac{|u_m(x)|}{2d(x, B_{m-1}))} \right) \left(\frac{u_m(x)}{|u_m(x)|} \frac{2CM^*(K)\sqrt{n}}{\sqrt{2^{m-1}}} \right). \end{aligned}$$

The Majorizing Measure Theorem tells us that the sum of the (left hand) coefficients in the infinite sum is less than 1, and therefore that for any linear operator S it is true that

$$|S(x)| \leq |S(v_{m_0}(x))| + \max_{m_0 < m < \infty} \frac{|S(u_m(x))|}{|u_m(x)|} \frac{2CM^*(K)\sqrt{n}}{\sqrt{2^{m-1}}}.$$

Hence

$$\operatorname{diam}(S(K)) \leq \max_{v \in B_{m_0}} |S(v)| + \sup_{m_0 < m < \infty, u_m \in A_m} \frac{|S(u_m)|}{|u_m|} \frac{2CM^*(K)\sqrt{n}}{\sqrt{2^{m-1}}}.$$

We will use Lemma 3.3.1 to show that for a random level k sign-projection the right hand side is less than $C_2\sqrt{\frac{k}{n}}$ with probability $1-2^{-c_3k}$, for any c_3, and a suitably chosen C_2 depending only on C_1 and c_3. We do this by showing that the following two events occur simultaneously with probability greater than $1-2^{-c_3k}$ (c_3 is any absolute constant)
(a) For every $v \in B_{m_0}$,

$$|S_k(v)| \le C_2'\sqrt{\frac{k}{n}}|v|$$

(b) For every $m > m_0$, for every $u_m \in A_m$,

$$|S_k(u_m)| \le C_2''\frac{\mathrm{diam}(K)\sqrt{2^{m-1}}}{M^*(K)\sqrt{n}}\sqrt{\frac{k}{n}}|u_m|.$$

Then we can simply take $C_2 = C_2' + 2CC_2''$.

Taking care of (a) is immediate, since the size of B_{m_0} is exponential in k. Thus we may choose in Lemma 3.3.1, say, $C_2' = t = 2 + c_3$, so that $\frac{1}{2}(t^2 - 1 - 2\log t) > 2 + 2c_3$ and get that for measure $1 - \exp(-k(2+2c_3))$ of the level k sign-projections, (a) holds true. Notice that in this part the interplay between the constants is straightforward: if we allow smaller c_3, we may choose a better (smaller) C_2', and C_1 does not yet enter into the picture at all.

For (b) we need indeed the isomorphic part of Lemma 3.3.1. For every m, the probability that (b) happens for this specific m is (denoting $d = \mathrm{diam}(K)$, and $M^* = M^*(K)$) at least

$$1 - 2^{2^{m+1}}\mathrm{Prob}[|S_k(x)| > t\sqrt{\frac{k}{n}}|x|],$$

for $t = \frac{d}{M^*}C_2''\frac{\sqrt{2^{m-1}}}{\sqrt{n}}$.

Since we consider $m > m_0$, we have that $2^{m-1} \ge k$, and hence $t \ge C_2''\frac{d}{M^*}\sqrt{\frac{k}{n}} \ge C_2''C_1$. Thus for $C_2''C_1$ big enough we, first of all, have $t > 1$ and we can apply Lemma 3.3.1. Secondly, in Lemma 3.3.1 for t which is big enough, one sees that the probability of failure is (for example) less than $e^{-t^2k/4}$. Thus the above expression, for a specific m and for C_2'' big enough (depending only on C_1), is at least

$$1 - 2^{2^{m+1}}e^{-\frac{k}{4}\frac{2^{m-1}C_2''^2}{n}(\frac{d}{M^*})^2}.$$

We are assuming that $kd^2/nM^{*2} \ge C_1^2$, which bounds this probability from below again by

$$1 - 2^{2^{m+1} - 2^{m-3}C_1^2C_2''^2/\ln 2}.$$

Making the choice $C_2''^2 = 16(2+c_3)\ln 2/C_1^2$, this probability is at least $1 - 2^{-2^{m+1}(1+c_3)}$. Adding up for all m the probability of failure in (b), and adding also the probability of failure in (a) as the first summand (where we have exchanged e by 2, only increasing the probability), we see that

$$\sum_{m=m_0}^{\infty} 2^{-2^{m+1}(c_3+1)} \leq 2^{-c_3 k},$$

which completes the proof. $\square$

We remark again that, as is seen in the proof, for every choice of the two constants c_3 and C_1, there exist a choice of $C_2(c_3, C_1)$ such that the conclusion of the theorem holds.

Secondly we remark again that this is a two sided inequality. Actually, one may even use Lemma 3.3.2 to see this: no smaller bound for the diameter shrinking in a random sign-projection can be obtained (up to the constants involved). Indeed, simply take the point with largest euclidean norm; under a random sign-projection the euclidean norm of this specific point is close to $\sqrt{\frac{k}{n}}$ times its original norm, and so the diameter is no smaller.

3.3.4 Some additional remarks

It seems appropriate to introduce the parameter $r^*(K)$ which is defined as

$$r^*(K) = \text{Ave}_{\varepsilon_i=\pm1} \sup\{\sum_i \varepsilon_i x_i : x \in K\}.$$

That is, we average the dual norm, not on the whole sphere as in the definition of $M^*(K)$, but only on the vertices of the cube. This parameter is known to be rather close to $\sqrt{n}M^*(K)$. Indeed, it is easy to show that there exists a universal constant C such that for every symmetric convex body K one has

$$\frac{1}{C} r^*(K) \leq \sqrt{n} M^*(K) \leq \sqrt{\log n}\, C r^*(K).$$

That is, up to a factor which is logarithmic in the dimension, the two parameters $r^*(K)$ and $\sqrt{n}M^*(K)$ are equivalent. Moreover, if the space whose unit ball is K° (the polar body of K) has some non-trivial cotype then (see [MaPi]) the two parameters $r^*(K)$ and $\sqrt{n}M^*(K)$ are equivalent up to a universal constant (for definitions and simple proofs see [Pis1]).

Another thing to notice is that we cannot expect the diameter of a random level k sign-projection to ever be smaller than $r^*(K)/\sqrt{n}$, since the diameter of a random level k sign-projection is clearly a decreasing function of k, and for $k = 1$ the average diameter is exactly $r^*(K)/\sqrt{n}$ (and therefore, by a standard concentration argument, this is true with high probability).

Let us discuss briefly the special example where the body K is the unit ball of ℓ_1^n, that is, $K = \mathrm{conv}\{\pm e_i\}$. In this case, a sign-projection of level k is simply the convex hull in $\mathbb{R}^k$ of the points

$$\{\pm(\xi_{1,i}, \xi_{2,i}, \ldots, \xi_{k,i}),\ i = 1 \ldots n\}.$$

That is, we pick n random vertices from the 2^k vertices of the k-dimensional cube (of edge length $2/\sqrt{n}$), and take their (absolute) convex hull. The diameter of this set is *exactly* $\sqrt{k/n}$ times the diameter of the original K, and this holds true for *every* k. Thus we have an example where the decay of diameter continues beyond the level $k^* = n(M^*)^2$ (which is in this case around $\log n$) and goes all the way to dimension $1 = r^{*2}$.

To conclude, we know how the diameter decreases for sign projections of levels greater than k^*. However, opposed to the orthogonal case in which we know that all projections to dimensions smaller than $\simeq k^*$ are euclidean balls of radius $\simeq M^*(K)$, in the sign-projections investigation we have no such natural reason for the decrease in diameter to stop, though we know that at dimension 1 it must be, with high probability, around $r^*/\sqrt{n}$, which is at least $M^*(K)/(C \log n)$.

3.3.5 Proof of inequality (3.3.3)

Let us show a proof of the fact that for $t_i = \pm 1$ (standard Bernoulli random variables), if $\sum {y_i}^2 = 1$ and $\lambda \in [0, \frac{1}{2})$, then

$$\mathbb{E}(\exp(\lambda(\sum y_i t_i)^2)) \le \frac{1}{\sqrt{1-2\lambda}} \tag{3.3.4}$$

(where $\mathbb{E}$ is taken with respect to the t_i's, that is, $\mathbb{E}(A) = \mathrm{Ave}_{t_i=\pm 1} A$). First notice that for any fixed t_i, $i = 1, \ldots, n$,

$$\exp(\lambda(\sum y_i t_i)^2) = \mathbb{E}_g \exp(\sqrt{2\lambda}(\sum y_i t_i) g)$$

where g is a standard Gaussian and $\mathbb{E}_g$ denotes expectation with respect to it. Using the independence we see that

$$\begin{aligned}
\mathbb{E}(\exp(\lambda(\sum y_i t_i)^2)) &= \mathbb{E}\mathbb{E}_g(\exp(\sqrt{2\lambda}(\sum y_i t_i))g) \\
&= \mathbb{E}_g \prod \mathbb{E}_{t_i}(\exp(\sqrt{2\lambda} y_i t_i) g) \\
&= \mathbb{E}_g \prod \cosh(\sqrt{2\lambda} y_i g) \\
&\le \mathbb{E}_g \prod \exp \lambda g^2 y_i^2 = \mathbb{E}_g \exp \lambda g^2 \sum y_i^2 \\
&= \frac{1}{\sqrt{1-2\lambda}}.
\end{aligned}$$

where we have used the easily verified inequality $\cosh x \leq e^{x^2/2}$. This completes the proof of the inequality. $\square$

3.4 Some further remarks on covering

3.4.1 Projections and Covering Numbers

In this section we present results concerning the behavior of covering numbers under projections. We show that the numbers $e_k(P_t K)$, the k^{th} entropy numbers of random rank t projections of a symmetric convex body K, exhibit a 'phase transition' phenomenon in some cases.

Below we use the standard jargon of the asymptotic theory of normed spaces, saying that a property is satisfied for a "random" projection of given rank k if it holds for a set of (orthogonal) rank k projections whose measure tends to 1 as the relevant parameters (k, n below) tend to infinity (where "measure" = "the normalized Haar measure on the corresponding Grassmann manifold").

We begin with a proposition which shows, as is typical in Asymptotic Geometric Analysis, a unified form of behavior for *all* dimensions and *all* convex bodies.

Proposition 3.4.1 *There exist universal constants c_1, C_1, and for every λ there exists a constant $C_2 = C_2(\lambda)$ depending only on λ such that, for any symmertic convex body $K \subset \mathbb{R}^n$ with $N(K, D) = 2^k$ and any integer t_0 with $k \leq t_0 \leq n$ we have*

(i) If $M^*(K \cap D) \leq \sqrt{\frac{t_0}{n}}$, then for every integer t with $t_0 \leq t \leq n$, the random rank t projection P_t satisfies

$$N(P_t K, C_1\sqrt{\frac{t}{n}}D) \leq 2^k \text{ and } N(P_t K, c_1\sqrt{\frac{t}{n}}D) \geq 2^k.$$

(ii) In the other direction, if the random rank t_0 projection P_{t_0} verifies

$$N(P_{t_0} K, \lambda\sqrt{\frac{t_0}{n}}D) \leq 2^k,$$

then necessarily $M^*(K \cap D) \leq C_2\sqrt{\frac{t_0}{n}}$ and, for any *integer t with $t_0 \leq t \leq n$, the random rank t projection P_t satisfies*

$$N(P_t K, C_2\sqrt{\frac{t}{n}}D) \leq 2^k \text{ and } N(P_t K, c_1\sqrt{\frac{t}{n}}D) \geq 2^k.$$

Proof For (i), denote $N(K, D) = 2^k$, assume that for some $t_0 \geq k$ we have $M^*(K \cap D) \leq \sqrt{\frac{t_0}{n}}$, and let t be some integer $t_0 \leq t \leq n$. Firstly, we can find a 1-separated set $\{x_i\}_{i=1}^{2^k}$ in K (see Fact 1.3.7 in Chapter 1). For a random rank-t projection P_t, with exponentially high probability, the projections $P_t x_i$ will form a $c\sqrt{t/n}$-separated set in $P_t K$, where c is a universal constant. This follows from the fact that $k \leq t$ and the results, for example, described in Section 3.2.6. Hence $N(P_t K, c_1 \sqrt{\frac{t}{n}} D) \geq 2^k$ for a universal c_1. Secondly, notice that we have $N(K, D) = 2^k$, and therefore (see Fact 1.3.6 in Chapter 1) $N(K, 2K \cap D) = 2^k$. For a random rank-t projection, since $M^*(2K \cap D) \leq 2\sqrt{\frac{t}{n}}$, we know that $P_t(2K \cap D)$ will be included in a ball of radius $C\sqrt{\frac{t}{n}}$, where C is a universal constant. This was used for example in the proof of 1.4.5 in Chapter 1, for reference see [Mi2] or Fact 3.1 in [MSz1]. So projecting the inclusion

$$K \subset \cup_{i=1}^{2^k} x_i + (2K \cap D_n)$$

we see that

$$P_t K \subset \cup_{i=1}^{2^k} P_t x_i + C\sqrt{\frac{t}{n}} D_k,$$

which implies $N(P_t K, C_1 \sqrt{\frac{t}{n}} D) \leq 2^k$, and completes the proof of (i).

To prove (ii), again denote $N(K, D) = 2^k$, assume that for some $t_0 \geq k$ we have that for a random rank t_0 projection P_{t_0} one has

$$N(P_{t_0} K, \lambda \sqrt{\frac{t_0}{n}} D) \leq 2^k.$$

(Notice that, by the same argument used in the proof of (i), for a random P_t one *always* has $N(P_t K, c_1 \sqrt{\frac{t}{n}} D) \geq 2^k$, hence it is not given as a condition).

Take a random rank t_0 projection, then for some set $\{y_i\}_{i=1}^{2^k}$ we have

$$P_{t_0} K \subset \cup_{i=1}^{2^k} y_i + \lambda \sqrt{\frac{t_0}{n}} D_{t_0}.$$

However, if $M^*(K \cap D) \geq C\sqrt{\frac{t_0}{n}}$ (for some universal C) then it is well known that a rank t_0 projection of $K \cap D$ will include a euclidean ball of radius $cM^*(K \cap D)$, where c is a universal constant. Therefore

$$cM^*(K \cap D) D_{t_0} \subset \cup_{i=1}^{2^k} P_{t_0} y_i + \lambda \sqrt{\frac{t_0}{n}} D_{t_0}.$$

Since $k < t_0$, this clearly implies (say, by volume estimates) that $M^*(K \cap D) < C(\lambda)\sqrt{\frac{t_0}{n}}$, as required. $\square$

Remark Notice that we could have actually added universal constants in many places in this proposition, to make it more general. We chose not to, in order to avoid notational inconveniences; the required changes in the proof would be minor.

We note that Proposition 1.4.8 implies, in the case when K is a convex hull of 2^k points, that the critical t_0 in the above proposition is bounded from above by $C_0 k(\log k)^6$, Notice that the validity of Conjecture 1.4.13 would imply that for this class of bodies in fact $t_0 \leq C_0 k$ for a universal C_0. Also, the main duality theorem of Chapter 1, Theorem 1.2.1, implies that Proposition 3.4.1 remains true if we replace the condition on $N(K, D)$ with a similar one on $N(D, K^\circ)$ (with additional universal constants). Similarly, we may replace the estimates on the behavior of covering numbers under projections with their dual analogues, describing the behavior of covering numbers under intersections with random subspaces.

Thus, the proposition is especially interesting in the case where the body K is the convex hull of 2^k points or less, and in many problems, when we encounter a body K with $N(K, D) = 2^k$, we lose no generality in assuming that our body is indeed such a polytope. We can reformulate the "Geometric Lemma" of Chapter 1, Section 1.4.3.4 in terms of behavior of covering numbers with projections. All this is summarized in the following proposition.

Proposition 3.4.2 *The four following statements are equivalent:*

(a) There exists $C > 0$ such that for all n, $K \subset \mathbb{R}^n$ and k,

$K = \mathrm{conv}\{x_i\}_{i \leq 2^k}$ and $N(K, D) \leq 2^k$ implies $M^*(K \cap D) \leq C\sqrt{\frac{k}{n}}$.

(b) There exists $C > 0$ such that for all n, $K \subset \mathbb{R}^n$ and k,

$K = \mathrm{conv}\{x_i\}_{i \leq 2^k}$ and $N(D, K^\circ) \leq 2^k$ implies $M^*(K \cap D) \leq C\sqrt{\frac{k}{n}}$.

(c) For every $c > 0$, there exists $C_1 > 0, c_1 > 0$ such that for all n, $K \subset \mathbb{R}^n$ and k,

$K = \mathrm{conv}\{x_i\}_{i \leq 2^k}$ and $N(K, D) \leq 2^k$ implies $N(P_{ck}K, C_1\sqrt{\frac{k}{n}}D_{ck}) \leq 2^{c_1 k}$,

for a random rank ck projection P_{ck}.

(d) For every $c > 0$, there exists $C_2 > 0, c_2 > 0$ such that for all n, $K \subset \mathbb{R}^n$ and k,

$K = \mathrm{conv}\{x_i\}_{i \leq 2^k}$ and $N(D, K^\circ) \leq 2^k$ implies $N(D_{E_{ck}}, C_2\sqrt{\frac{k}{n}}K^\circ \cap E_{ck}) \leq 2^{c_2 k}$,

for a random ck-dimensional subspace E_{ck}.

Proof The equivalence between the first two, which are the Geometric Lemma and the Dual Geometric Lemma, was shown in Chapter 1, Section 1.4.3.4. Proofs of all other implications follow the same lines as that of Proposition 3.4.1 above.

We want to emphasize that the result above (in the form of Proposition 3.4.1) is an example for a phase transition behavior of the function $e_k(P_tK)$. Recall the definition of entropy numbers: the number $e_k(P_tK)$ is the infimal ε such that for a random rank-t projection, $N(P_tK, \varepsilon D_t) \le 2^{k-1}$, with probability, say, greater than $9/10$. Thus $e_1(P_tK)$ is simply the (expected) diameter of the set P_tK. The behavior of $e_1(P_tK)$ is well known. We denote by d the diameter of K (that is, $e_1(P_nK)$), and by M^*, as usual, half the mean width of the body K (hence, with high probability on P_1 which are one dimensional projections, $e_1(P_1K) = M^*$). Denote also $k^* = (M^*/d)^2 n$, so that $\sqrt{k^*/n} = M^*/d$. Then, for $t \gtrsim k^*$, with very high probability on the choice of a random projection P_t, one has $e_1(P_tK) \simeq \sqrt{t/n}d$, and for $t \lesssim k^*$, with very large probability on the choice of a random projection P_t, one has $e_1(P_tK) \simeq M^*$. Thus, we have a phase transition of the function $e_1(P_t)$, and the location of this phase transition depends on the parameters of the body K we are projecting, namely its diameter and its mean width. Notice, still, that we have been somewhat imprecise, since there are always constants involved. The probability we encounter in fact gets closer to 1 the farther we are from the critical k^*; In other words, at k^* and next to it we know the result only up to universal constants (as stated above) whereas once we are out of the 'dangerous zone' close to k^* we actually know the behavior of e_1 isometrically.

How about the other functions e_k? Proposition 3.4.1 (and a simple extension) delivers the required information, which is not complete. As a first example conside the body $K = (1 + \frac{k}{n})D_n$; thus $N(K, D) \simeq 2^k$ and we have $e_k(n) = 1$. It is easy to check that for this body t_0 of Proposition 3.4.1 equals n. All the projections are again euclidean balls, and so $e_k(t) = 1$ for $t > k$, and it is not hard to check that, since $N(D_t, \delta D_t) \simeq e^{t \log 1/\delta}$, for $t < k$ one has $e_k(t) \simeq e^{-k/t}$. So, even in this simple example, we see a phase transition at k. Another simple example is when K is a segment. We take $K = [0, 2^k]$, so that $e_k(n) = 1$. Since with high probablity $P_t(K)$ is a segment of length $\sqrt{t/n}2^k$, we see that $e_k(t) \simeq \sqrt{\frac{t}{n}}$ (and t_0 in this case is 1): no phase transition.

What can we say in the general case? It is not difficult to see, from considerations such as the ones in Proposition 3.4.1 that, denoting $t_0 = (M^*(2e_k(n)^{-1}K\cap$

$D)^2n$,

- For $t > ck$ we always have $e_k(t) \geq C(c)\sqrt{t/n}$.
- For $t > \max\{k, t_0\}$ we have $e_k(t) \simeq \sqrt{t/n}$.
- For $k < t < t_0$ we have $e_k(t) \geq C(c)\sqrt{t_0/n}$.
- For $t < k^* < k$ we have $e_k(t) \simeq e^{-t/n}M^*(K)$.

Thus, for example, in the case where $k << t_0$ we always have a phase transition at t_0. In the case $k < ck^*$ we have a complete picture of $e_k(t)$ as a function of t, with two phase transitions at t_0 and at k (notice that *always* $k^* \leq 4t_0$). However, in the case $k^* < t_0 < k$ for example, we do not know how $e_k(t)$ will behave in the region $t \in (k^*, k)$, and further investigation is required in order to have the complete picture (and that picture might depend on more parameters).

3.4.2 Weighted Covering

In this section we define a natural extension of the notion of covering numbers, which is covering numbers with weights. We then show that up to universal constants, these two notions are equivalent.

Let K, T be convex centrally symmetric bodies in $\mathbb{R}^n$. Recall, once again, the standard definition of covering numbers:

$$N(K,T) = \min\{N : \exists x_1 \ldots x_N \in \mathbb{R}^n,\ \forall x \in K\ \exists i\ \ x \in x_i + T\}.$$

It can be re-written in term of densities in the following way

$$N(K,T) = \min\{\int w : w = \sum_i \delta_{x_i},\ \forall x \in K \quad \int_{x+T} w \geq 1\}.$$

It seems natural to define a covering by-two-layers number

$$N_2'(K,T) = \min\{N : \exists x_1 \ldots x_N \in \mathbb{R}^n,\ \forall x \in K\ \exists i \neq j\ \ x \in (x_i+T)\cap(x_j+T)\}.$$

And the normalized $N_2(K,T) = \frac{1}{2}N_2'(K,T)$. Or, in the other form of writing (already normalized)

$$N_2(K,T) = \min\{\int w : w = \sum_i \frac{1}{2}\delta_{x_i},\ \forall x \in K \quad \int_{x+T} w \geq 1\}.$$

Of course, $N_2(K,T) \leq N(K,T)$ because one could always choose the same density w as before, this time considering each δ_{x_i} as $\frac{1}{2}\delta_{x_i} + \frac{1}{2}\delta_{x_i}$, but the minimum might be attained on w which assigns weight $1/2$ to some point.

In a similar way one can define

$$N_k(K,T) = \min\{\int w : w = \sum_i \frac{1}{k}\delta_{x_i},\ \forall x \in K \quad \int_{x+T} w \geq 1\},$$

and it is clear that $N_k(K,T) \leq \frac{k-l}{k} N_{k-l}(K,T) + \frac{l}{k} N_l(K,T)$. Passing to continuous densities, we can define

$$N_\infty(K,T) = \inf\{\int w : w \geq 0,\ \forall x \in K \quad \int_{x+T} w \geq 1\}. \tag{3.4.1}$$

Notice that $N_\infty(K,T) \geq \frac{\mathrm{Vol}(K)}{\mathrm{Vol}(T)}$; Indeed,

$$\mathrm{Vol}(K) = \int_K 1 \leq \int_K \int_{x+T} w \leq \int_{\mathbb{R}^n} \int_{x+T} w = \mathrm{Vol}(T) \int w.$$

We can also define a similar notion of separation. The usual maximal separated set's cardinality can be written in the following form

$$M(K,T) = \max\{\int w : w = \sum_i \delta_{x_i},\ \forall x \in K \quad \int_{x+\frac{1}{2}T} w \leq 1\}.$$

It is easy to check that $N(K,T) \leq M(K,T)$, simply because if the maximum is attained on a certain w, then surely $w + \delta_x$ cannot be permissible, hence there is some $y \in K$ for which $\int_{y+\frac{1}{2}T}(w + \delta_x) > 1$ and this means precisely that $x \in y + \frac{1}{2}T$. Moreover, since we are dealing with delta functions, it must be the case that $\int_{y+\frac{1}{2}T} w = 1$. Since $x \in y + \frac{1}{2}T$ implies that $y + \frac{1}{2}T \subset x + T$, we are now sure that $\int_{x+T} w \geq 1$. This may seem like a complicated proof of a trivial fact (a maximal separated set is always a covering) but the point is it translates easily to the continuous case. First let us define the sequence of layer-separated sets

$$M_k(K,T) = \max\{\int w : w = \sum_i \frac{1}{k}\delta_{x_i},\ \forall x \in K \quad \int_{x+\frac{1}{2}T} w \leq 1\}.$$

This means precisely that every point is covered by at most k layers (each of "weight" $\frac{1}{k}$). As in coverings, $M_k(K,T) \geq \frac{l}{k} M_l(K,T) + \frac{k-l}{k} M_{k-l}(K,T)$, since there are more candidates participating in the maximum.

In the same way as above we can show that $M_k(K,T) \leq N_k(K,T)$: Assume the maximum is attained by some specific w, and let $x \in K$, then $w + \frac{1}{k}\delta_x$ is not permissible. Hence there exists $y \in K$ with $\int_{y+\frac{1}{2}T}(w + \frac{1}{k}\delta_x) > 1$. Since all these integrals are multiplies of $\frac{1}{k}$, it must be that $\int_{y+\frac{1}{2}T} w = 1$, and that $x \in y + \frac{1}{2}T$. This implies, since $y + \frac{1}{2}T \subset x + T$, that $\int_{x+T} w \geq 1$.

Of course, we didn't really need to prove this, because

$$M_k(K,T) \geq M(K,T) \geq N(K,T) \geq N_k(K,T),$$

however, in this way we showed that every maximal layer-separated density is itself a layer-covering.

As for the continuous version, we define

$$M_\infty(K,T) = \sup\{\int w : w \geq 0,\ \forall x \in K \quad \int_{x+\frac{1}{2}T} w \leq 1\}.$$

Notice that $M_\infty(K,T) \leq \frac{\mathrm{Vol}(K+\frac{1}{2}T)}{\mathrm{Vol}(\frac{1}{2}T)}$; Indeed, one permissible density is simply $w = \frac{1}{\mathrm{Vol}(\frac{1}{2}T)}\mathbf{1}_{K+\frac{1}{2}T}$.

Again, every continuous separated density is automatically a covering density since if the supremum is (almost, say $-\varepsilon/2$) attained on w, then $w + \varepsilon\delta_x$ is not permissable, which in turn gives that for some $y \in x + \frac{1}{2}T$ $\int_{y+\frac{1}{2}T} w > 1 - \varepsilon$ and hence $\int_{x+T} w > 1 - \varepsilon$. This is true for any ε, hence this integral is ≥ 1. We have thus shown that $N_\infty(K,T) \leq M_\infty(K,T)$, although we knew it already, as above.

It is also trivial to see that $M(K,T) \leq N(K, \frac{1}{2}T)$, since every copy of $\frac{1}{2}T$ can include at most one of the separated points. We can also show that

$$M_\infty(K,T) \leq N_\infty(K, \frac{1}{2}T).$$

Indeed, what we want to show is that if w_1 is such that for every $x \in K$ one has $\int_{x+\frac{1}{2}T} w_1 \leq 1$ and w_2 is such that for every $x \in K$ one has $\int_{x+\frac{1}{2}T} w_2 \geq 1$, then (naturally) $\int w_1 \leq \int w_2$. This is true since

$$\begin{aligned}\int_{R^n} w_1(x)dx &\leq \int_{R^n} w_1(x) \int_{x+\frac{T}{2}} w_2(y)dydx \\ &= \int_{R^n} w_2(y) \int_{y+\frac{T}{2}} w_1(x)dxdy \leq \int_{R^n} w_2(y)dy.\end{aligned}$$

In conclusion, we see that these numbers are equivalent, up to a constant 2, to the usual covering numbers.

$$N(K,T) \leq M(K,T) \leq M_\infty(K,T) \leq N_\infty(K, \frac{1}{2}T) \leq N(K, \frac{1}{2}T)$$

The advantage of weighted covering numbers, is, first of all, that at least ideologically the should be easier to compte. Indeed, this should be compared with the difference between solving a linear programming problem with integers or with reals. Of course, definition (3.4.1) is exactly that of a linear optimization problem, albeit infinite dimensional. More down to earth advantages can seen by the following example: when you try covering the euclidean ball with small copies of it, you need to find a good distribution of point densities. If you are working, instead, with weighted coverings, first notice that an average of coverings is again a covering, and in the case of a ball you can rotate a covering as well. Thus, you may assume your covering weight function w is radial and optimize on a *one variable* function. The above reasoning tells you that up to a constant the answers will be the same (of course, in this case almost everything is known). More generally, when one works with weighted covers it is easoer to invoke the symmetry of the sets to simplify the optimization problem.

3.4.3 Remark - No duality with rotations.

To end this chapter (and thesis), we would like to make a simple observation. Many of the methods for working with covering numbers are invariant under rotations. For example, if we bound from above the covering number of K by T by means of volume, or by Sudakov's inequality, the same bounds are applicable if you rotate one of the bodies. Further, they are applicable if you allow to cover with rotated copies of T, that is, if you ask for the minimal number N such that there exist centers x_i and orthogonal rotations u_i so that K is included in $\cup_{i=1}^{N} x_i + u_i T$. However, this number can clearly be much smaller than the covering number.

A natural question would then be whether a duality theorem can exist for this number. The answer is no, and this can be seen from the simple example of covering the euclidean ball with a round cylinder. Indeed, denote, for $i < j < k$,

$$S_{i,j,k} = \{x \in \mathbb{R}^n : x_i^2 + x_j^2 + x_k^2 \le \frac{3}{n}\}.$$

The sets $S_{i,j,k}$ are rotations of the cylinder

$$T = \{x \in \mathbb{R}^n : x_1^2 + x_2^2 + x_3^2 \le \frac{3}{n}\}.$$

For simplicity assume that $n = 3m$, and notice that

$$D \subset \cup_{i=1}^{m} S_{3i-2,3i-1,3i},$$

because for every point $x \in D$

$$\sum_{i=1}^{m} (x_{3i-2}^2 + x_{3i-1}^2 + x_{3i}^2) \le 1.$$

Thus, with m rotations of T we can cover D. However, T° is a three dimensional euclidean ball with radius $\sqrt{\frac{n}{3}}$, and to cover it with euclidean unit balls one would need $\simeq n^{3/2}$ copies (with euclidean balls rotations do not matter). This shows that the two numbers are different. Doing the same trick with k instead of 3 would give one covering $\simeq \frac{n}{k}$ and the other one $\simeq (\frac{n}{k})^{k/2}$, so that for $k = \sqrt{n}$ say, we compare $\sqrt{n}$ with $e^{\sqrt{n}\log n}$.

Bibliography

[Ach] D. Achlioptas, *Database-friendly Random Projections.* Proceedings of PODS 01, 274-281.

[AlDi] D. Alfers and H. Dinges, *A Normal Approximation for Beta and Gamma Tail Probabilities*, Wahrscheinlichkeitstheorie 65 (1984), 399-420.

[Art1] S. Artstein, *Proportional concentration phenomena on the sphere.* Israel J. Math. 132 (2002), 337–358.

[Art2] S. Artstein, *The Change in the Diameter of a Convex Body Under a Random Sign-Projection*, to appear in GAFA Seminar Notes, Springer Lect. Notes series.

[ABBN1] S. Artstein, K. Ball, F. Barthe and A. Naor, *Solution of Shannon's Problem on the Monotonicity of Entropy*, to appear in Journal of Amer. Math. Soc. (2004).

[ABBN2] S. Artstein, K. Ball, F. Barthe and A. Naor, *On the rate of convergence in the entropic central limit theorem*, to appear in Probability Theory and Related Fields (2004).

[AMS1] S. Artstein, V.D. Milman and S.J. Szarek, *More on the duality conjecture for entropy numbers.* C. R. Math. Acad. Sci. Paris 336 (2003), no. 6, 479–482.

[AMS2] S. Artstein, V.D. Milman and S.J. Szarek, *Duality of metric entropy in Euclidean space.* C. R. Math. Acad. Sci. Paris 337 (2003), no. 11, 711-714.

[AMS3] S. Artstein, V.D. Milman and S.J. Szarek, *Duality of Metric Entropy*, to appear in Annals of Mathematics (2004).

[AMST] S. Artstein, V.D. Milman, S.J. Szarek and N. Tomczak-Jaegermann, *Convex Separation and Duality*, to appear in GAFA Journal.

[BaEm] D. Bakry and M. Emery. *Diffusions hypercontractives*, Séminaire de Probabilités XIX, number 1123 in Lect. Notes in Math., Springer (1985) 179–206.

[BBN] K. Ball, F. Barthe and A. Naor, *Entropy jumps in the presence of a spectral gap*, Duke Math. J. 119 (2003), no. 1, 41–63.

[Bar] A.R. Barron, *Entropy and the central limit theorem*, Ann. Probab. 14 (1986), 336–342.

[BaJo] A.R. Barron and O. Johnson, *Fisher information inequalities and the central limit theorem* Preprint, ArXiv:math.PR/0111020.

[Bla] N.M. Blachman, *The convolution inequality for entropy powers*, IEEE Trans. Info. Theory 2 (1965), 267–271.

[BPST] J. Bourgain, A. Pajor, S.J. Szarek and N. Tomczak-Jaegermann, *On the duality problem for entropy numbers of operators.* Geometric aspects of functional analysis (1987–88), Lecture Notes in Math., 1376, Springer, Berlin-New York (1989) 50–63.

[Bro] L.D. Brown, *A proof of the central limit theorem motivated by the Cramer-Rao inequality*, Kalliampur et al., editor, Statistics and Probability: Essays in Honor of C. R. Rao,, Amsterdam, 1982. North-Holland, 314–328.

[CaSo] E.A. Carlen and A. Soffer, *Entropy production by block variable summation and central limit theorems*, Commun. Math. Phys., 140 no.2 (1991), 339–371.

[Csi] I. Csiszar, *Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen*, Publications of the Mathematical Institute, Hungarian Academy of Sciences, VII, Series A (1962), 137–157.

[DaSz] K.R. Davidson and S.J. Szarek, *Local operator theory, random matrices and Banach Spaces*, Handbook of the geometry of Banach spaces, Vol I, North Holland, Amsterdam (2001), 317–366.

[DiFr] P. Diaconis and D. Freedman, *A dozen de Finetti-style results in search of a theory*, Ann. Inst. Henri Poincaré, 23 (1987), 397–423.

[Du] R.M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes.* J. Funct. Anal. 1 (1967), 290–330.

[GiMi] A. Giannopoulos and V.D. Milman, *Euclidean structure in finite dimensional normed spaces*, Handbook of the Geometry of Banach Spaces, Vol. I, North-Holland, Amsterdam (2001), 707–779.

[GKS] Y. Gordon, H. König and C. Schütt, *Geometric and probabilistic estimates for entropy and approximation numbers of operators.* J. Approx. Theory 49 (1987), no. 3, 219–239.

[Gr] M. Gromov, *Isoperimetry of waists and concentration of maps* Geometric and Functional Analysis, 13, no. 1 (2003) 178–215.

[GLS] M. Grötschel, L. Lovász and A. Schrijver, *Geometric algorithms and combinatorial optimization.* Algorithms and Combinatorics, 2. Springer-Verlag, Berlin, 1993.

[HJ] J. Hoffmann-Jørgensen, *Sums of independent Banach space valued reandom variables*, Aarhus Univ. Preprint series 15 (1972/2).

[JoLi] W.B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math. 26 (1984), pp. 189–206.

[KöM] H. König and V.D. Milman, *On the covering numbers of convex bodies*, Geometric aspects of functional analysis (1985–86), Lecture Notes in Math., vol. 1267, Springer, Berlin-New York (1987) 82–95.

[KuLi] J. Kuelbs and W.V. Li, *Metric entropy and the small ball problem for Gaussian measures.* J. Funct. Anal. 116 (1993), no. 1, 133–157.

[Kul] S. Kullback, *A lower bound for discrimination information in terms of variation*, IEEE Trans. Info. Theory, 4 (1967), 126–127.

[Led] M. Ledoux, *Isoperimetry and Gaussian analysis*, Lectures on probability theory and statistics (Saint-Flour, 1994), Lecture Notes in Math. 1648, Springer, Berlin, (1996), 165–294.

[LiLi] W.V. Li and W. Linde, *Small deviations of stable processes via metric entropy.* J. Theor. Probab., to appear.

[Lie] E.H. Lieb, *Proof of an entropy conjecture of Wehrl*, Comm. Math. Phys. 62, no. 1 (1978), 35–41.

[Lin] Ju.V. Linnik, *An information theoretic proof of the central limit theorem with lindeberg conditions*, Theory Probab. Appl. 4 (1959), 288–299.

[Mau] B. Maurey, *Théorèmes de factorisation pour les opérateurs linéairs à valeurs dans les espaces L^p*, Astérisque 11, 1–163.

[MaPi] B. Maurey and G. Pisier, *Séries de variables aleatoires vectorielles independantes et propriétés geometriques des espaces de Banach*, Studia Math. 58, 45–90.

[Mi1] V.D. Milman, *A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies* Funkcional. Anal. i Prilov zen. 5 (1971) no. 4 28–37.

[Mi2] V.D. Milman, *A note on a low M^*-estimate*, Geometry of Banach Spaces (Strobl, 1989), London Math. Soc. Lecture notes Ser., 158, Cambridge Univ. Press, Cambridge (1990) 219–229.

[MPa] V.D. Milman and A. Pajor, *Entropy and Asymptotic Geometry of Non-Symmetric Convex Bodies*, Advances in Mathematics 152 (2000), 314–335.

[MSch] V.D. Milman and G. Schechtmann, *Asymptotic Theory of Finite Dimensional Normed Spaces*, Lecture Notes in Mathematics 1200, Springer-Verlag, Berlin, 1986.

[MSz1] V.D. Milman and S.J. Szarek, *A geometric lemma and duality of entropy numbers*, Geometric aspects of functional analysis (1996–2000), Lecture Notes in Math., vol. 1745, Springer, Berlin-New York (2000) 191–222.

[MSz2] V.D. Milman and S.J. Szarek, *A geometric approach to duality of metric entropy*, C. R. Acad. Sci. Paris Ser. I Math. 332 (2001) no. 2, 157–162.

[PaTo1] A. Pajor and N. Tomczak-Jaegermann, *Subspaces of small codimension of finite dimensional Banach spaces*, Proc. Amer. Math. Soc. 97 (1986), 637–642.

[PaTo2] A. Pajor and N. Tomczak-Jaegermann, *Volume ratio and other s-numbers of operators related to local properties of Banach spaces*, J. Func. Anal. 87 (2) (1989), 273–293.

[Pie] A. Pietsch, *Theorie der Operatorenideale* (Zusammenfassung), Friedrich-Schiller-Universität Jena, 1972.

[Pin] M.S. Pinsker, *Information and information stability of random variables and processes*, Holden-Day, San Francisco, 1964.

[Pis1] G. Pisier, *Probabilistic Methods in the Geometry of Banach Spaces*, Probability and Analysis (Varenna, 1985) Lecture Notes in Math., 1206, Springer, Berlin, 1986, 167–241.

[Pis2] G. Pisier, *A new approach to several results of V. Milman*, J. Reine Angew. Math. 393 (1989), 115–131.

[Pis3] G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge Tracts in Mathematics, 94, Cambridge University Press, Cambridge (1989).

[Pis4] G. Pisier, *Remarques sur un résultat non publié de B. Maurey*, Seminar on Functional Analysis, 1980–1981, Exp. No. V, 13 pp., École Poltech., Palaiseau, 1981.

[Sch] C. Schütt, *Entropy numbers of diagonal operators between symmetric Banach spaces.* J. Approx. Theory 40 (1984), no. 2, 121–128.

[ShWe] C.E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, IL, 1949.

[Shi] R. Shimizu. *On Fisher's amount of information for location family*, Patil et al., editor, A modern course on statistical distributions in scientific work, Boston, MA, 1974. D. Reidel.

[Sta] A.J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Info. Control, 2 (1959), 101–112.

[Ta1] M. Talagrand, *Majorizing Measures without Measures.* Ann. Probab. 29 (1), (2001), 411–417.

[Ta2] M. Talagrand, *The Generic Chaining.* (To appear in Springer Verlag Publications).

[Ta3] M. Talagrand, *The Sherington-Kirkpatric Model: a Challenge for Mathematicians*, Probability Theory and Related Fields 110, 1998, 109–176.

[Tom] N. Tomczak-Jaegermann, *Dualité des nombres d'entropie pour des opérateurs à valeurs dans un espace de Hilbert*, C. R. Acad. Sci. Paris Sér. I Math. 305 (1987), no. 7, 299–301.

[Va] S.R.S. Varadhan, *Large deviations and applications*, SIAM Publications, Philadelphia, 1984.