Lectures in Geometric Functional Analysis

Roman Vershynin

Contents

Check the numbering of figures

CHAPTER 1

Functional analysis and convex geometry

Geometric functional analysis studies high dimensional linear structures. Some examples of such structures are Euclidean and Banach spaces, convex sets and linear operators in high dimensions. A central question of geometric functional analysis is: what do typical n-dimensional structures look like when n grows to infinity? One of the main tools of geometric functional analysis is the theory of concentration of measure, which offers a geometric view on the limit theorems of probability theory. Geometric functional analysis thus bridges three areas – functional analysis, convex geometry and probability theory. The course is a systematic introduction to the main techniques and results of geometric functional analysis.

1. Preliminaries on Banach spaces and linear operators

We begin by briefly recalling some basic notions of functional analysis. A *norm* defined on a linear vector space X is a function $\|\cdot\| : X \to \mathbb{R}$ that satisfies

(1) nonnegativity: $\|x\| \geq 0$ for all $x \in X$ and $\|x\| = 0$ iff $x = 0$;
(2) homogeneity: $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbb{R}$ and for all $x \in X$;
(3) triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$.

A *Banach space* is a complete normed space. We now recall some examples of classical Banach spaces.

Examples 1.1. 1. *The space of continuous functions* $C[0,1]$ consists of the functions $f : [0,1] \to \mathbb{R}$ that are continuous. It is a Banach space with respect to the sup-norm

$$\|f\|_\infty = \sup_{t \in [0,1]} |f(t)|.$$

2. For $1 \leq p < \infty$, *the space of p-integrable functions* $L_p[0,1]$ consists of the functions $f : [0,1] \to \mathbb{R}$ such that $|f|^p$ is Lebesgue integrable on $[0,1]$. It is a Banach space with respect to the norm

$$\|f\|_p = \Big(\int_0^1 |f(t)|^p \, dt \Big)^{1/p}.$$

3. *The space of essentially bounded functions* $L_\infty[0,1]$ consists of all Lebesgue measurable functions $f : [0,1] \to \mathbb{R}$ such that

$$\|f\|_\infty = \operatorname*{esssup}_{t \in [0,1]} |f(t)| = \inf_{f = g \text{ a.e.}} \sup_{t \in [0,1]} |g(t)| < \infty.$$

It is a Banach space with respect to this norm.

4. For $1 \le p < \infty$, *The space of p-summable sequences* ℓ_p consists of all sequences of real numbers $x = (x_i)$ such that the series $\sum_i |x_i|^p$ converges. It is a Banach space with respect to the norm

$$\|x\|_p = \Big(\sum_{i=1}^{\infty} |x_i|^p\Big)^{1/p}.$$

5. *The space of bounded sequences* ℓ_∞ consists of all sequences of real numbers $x = (x_i)$ such that

$$\|x\|_\infty = \sup_{i\in\mathbb{N}} |x_i|.$$

It is a Banach space with respect to this norm.

6. Every Hilbert space H is a Banach space with the norm $\|x\| = \sqrt{\langle x, x\rangle}$, where $\langle\cdot,\cdot\rangle$ denotes the inner product on H.

Let X, Y be Banach spaces. A *linear operator* is a map $T : X \to Y$ that satisfies $T(ax + by) = aT(x) + bT(y)$ for all $x, y \in X$ and $a, b \in \mathbb{R}$. In particular, if $Y = \mathbb{R}$, then T is called a *linear functional.* A linear operator $T : X \to Y$ is *bounded* if there exists some $M > 0$ such that $\|Tx\| \le M\|x\|$ for every $x \in X$. The infimum of such numbers M is called the *norm of* T:

$$\|T\| = \sup_{x\in X\setminus\{0\}} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\|.$$

A linear operator T is continuous if and only if it is bounded.

A linear operator $T : X \to Y$ is called an *isomorphism* if it is bijective and both T and T^{-1} are bounded. The Open Mapping Theorem actually implies that if T is bijective and bounded then so it T^{-1}. An isomorphism T is called an *isometry* if $\|T\| = \|T^{-1}\| = 1$, or equivalently if $\|Tx\| = \|x\|$ for all $x \in X$. Banach spaces X and Y are called isomorphic (resp. isometric) if there exists an isomorphism (resp. isometry) $T : X \to Y$. Any two finite dimensional Banach spaces of the same dimension are isomorphic, but not necessarily isometric.

Many arguments in functional analysis are based on duality. It provides a convenient tool: when two objects are in dual relation, one can study one of them and deduce properties of the other one using duality. If X is a normed space, then its *dual space* X^* consists of all continuous linear functionals on X, equipped with the operator norm. The dual space X^* is always a Banach space. The dual spaces of some classical spaces can be described via the known representation theorems.

EXAMPLES 1.2. 1. $(L_p[0,1])^* = L_q[0,1]$ for $p, q \in (1,\infty)$ such that $1/p + 1/q = 1$, and also for $p = 1$, $q = \infty$. Specifically, every continuous linear functional $F \in (L_p[0,1])^*$ has the form

$$F(f) = \int_0^1 f(t)g(t)\,dt, \quad f \in L_p[0,1]$$

for some function $g \in L_q[0,1]$.

2. $(\ell_p)^* = \ell_q$ for the same pairs p, q as in the previous example. Specifically, every continuous linear functional $F \in (\ell_p)^*$ has the form

$$F(x) = \sum_{i=1}^{\infty} x_i y_i, \quad x \in L_p[0,1]$$

for some function $y \in L_q[0,1]$.

3. Riesz representation theorem states for every Hilbert space H one has $H^* = H$. Specifically, every continuous linear functional $F \in H^*$ has the form

$$F(x) = \langle x, y \rangle, \quad x \in H$$

for some vector $y \in H$.

Duality is also defined for linear operators $T : X \to Y$. For a linear operator $T : X \to Y$, the *adjoint operator* is a linear operator $T^* : Y^* \to X^*$ defined by $(T^* f)(x) = f(Tx)$ for all $f \in Y^*$ and $x \in X$. If T is bounded then so is T^*, and we have $\|T^*\| = \|T\|$.

There are two natural classes of Banach spaces associated with a Banach space X: closed linear *subspaces* $E \subseteq X$ (with the norm induced from X) and *quotient spaces* X/E. The quotient space is defined using the following equivalence relation on X: $x \sim y$ iff $x - y \in E$. The quotient space X/E then consists of the equivalence classes

$$[x] = \{y \in X : y \sim x\} = \{x + e : e \in E\}.$$

X/E is a Banach space equipped with the norm

$$\|[x]\| = \inf_{y \in [x]} \|y\| = \mathrm{dist}_X(0, [x]) = \inf_{e \in E} \|x - e\|.$$

Figure 1 illustrates the concept of quotient space.

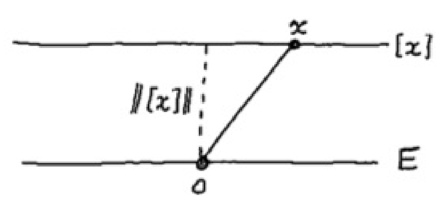

FIGURE 1. Quotient space X/E

Subspaces and quotient spaces are in duality relation. Recall that the *annihilator* of $E \subset X$ is a subspace in X^* defined as $E^\perp = \{f \in X^* : f(e) = 0 \quad \forall e \in E\}$. The next proposition says that the dual of a subspace is isometric to a quotient space (of the dual space), and the dual of a quotient space is isometric to a subspace (of the dual space).

Proposition 1.3 (Duality between subspaces and quotients). *Let $E \subset X$ be a linear subspace. Then*

(i) E^* is isometric to $X^*/E^\perp$;

(ii) $(X/E)^*$ is isometric to $E^\perp$.

For the proof, see [**?**, Exercise 9.10.25].

As we said, geometric functional analysis studies finite-dimensional normed spaces. Such spaces are always complete, and all their linear subspaces are closed. All n-dimensional Banach spaces are isomorphic to each other (but not isometric!) All linear operators on such spaces are bounded.

Classical examples of finite dimenisonal spaces are the n-dimensional versions of the spaces ℓ_p, which are denoted by ℓ_p^n. Thus ℓ_p^n is the linear vector space $\mathbb{R}^n$ equipped with the norm

$$\|x\|_p = \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p} \quad \text{if } 1 \le p < \infty; \qquad \|x\|_\infty = \max_{i \le n} |x_i|.$$

2. A correspondence between Banach spaces and convex bodies

A set K in $\mathbb{R}^n$ is called *convex* if it contains the interval joining any two points in K. A vector x is called a *convex combination* of points $x_1, \dots, x_m$ in $\mathbb{R}^n$ if it has the form

$$x = \sum_{i=1}^m \lambda_i x_i \quad \text{where } \lambda_i \ge 0, \quad \sum_{i=1}^m \lambda_i = 1.$$

A set K is convex if and only if it contains all convex combinations of any set of points from K.

Convex sets can be constructed from arbitrary sets by taking convex hulls. The *convex hull* of a set A in $\mathbb{R}^n$ is the minimal convex set containing A, denoted $\mathrm{conv}(A)$. Equivalently, $\mathrm{conv}(A)$ is the set of all convex combinations of points from A.

Recall that a convex set K in $\mathbb{R}^n$ is called *symmetric* if $K = -K$ (i.e. $x \in K$ implies $-x \in K$). A bounded convex set with nonempty interior is called a *convex body*.

Let X be an n-dimensional Banach space. Choosing a bijective map $X \to \mathbb{R}^n$, we can identify X with the linear vector space $\mathbb{R}^n$ equipped with some norm $\|\cdot\|$, thus often writing

$$X = (\mathbb{R}^n, \|\cdot\|)$$

for an arbitrary n-dimensional Banach space X. The following simple proposition establishes a correspondence between finite-dimensional Banach spaces and symmetric convex bodies in $\mathbb{R}^n$.

Proposition 2.1 (Banach spaces and symmetric convex bodies). *1. Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a Banach space. Then its unit ball B_X is a symmetric convex body in $\mathbb{R}^n$.*

2. Let K be a symmetric convex body in $\mathbb{R}^n$. Then K is the unit ball of some normed space $X = (\mathbb{R}^n, \|\cdot\|)$. Indeed, consider the Minkowski functional *of K defined for all $x \in \mathbb{R}^n$ as*

$$\|x\|_K = \inf\left\{t > 0 : \frac{x}{t} \in K\right\}.$$

Then $\|\cdot\|_K$ defines a norm on $\mathbb{R}^n$, and K is the unit ball of the Banach space $(\mathbb{R}^n, \|\cdot\|_K)$.

proof?

Figure 2 illustrates the definition of Minkowski functional.

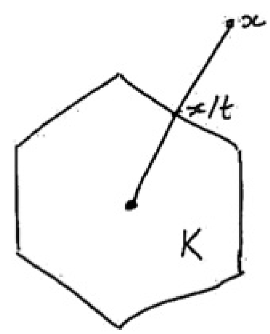

FIGURE 2. Minkowski functional

The correspondence between Banach spaces and convex bodies established in Proposition 2.1 is a simple but extremely useful tool. It allows one to use arguments from convex geometry in functional analysis and vice versa. This correspondence is especially useful for the classical spaces ℓ_p^n. Their unit balls, denoted by B_p^n, are easy to describe geometrically.

EXAMPLES 2.2. 1. The unit ball of the n-dimensional Euclidean space ℓ_2^n is the unit Euclidean ball in $\mathbb{R}^n$.

2. The unit ball of ℓ_1^n is the so-called cross-polytope in $\mathbb{R}^n$. The cross-polytope is the symmetric convex hull of the canonical basis vectors e_i in $\mathbb{R}^n$ and their opposites: $B_1^n = \mathrm{conv}(\pm e_i : \ i = 1, \dots, n)$.

3. The unit ball B_∞^n of ℓ_∞^n is the cube in $\mathbb{R}^n$ with center at the origin and side 2.

Figure 2 illustrates the shapes of the balls B_1^n, B_2^n and B_∞^n in dimension $n = 2$.

The concepts of subspaces and quotient spaces can also be interpreted in the language of convex geometry. Using Proposition 2.1, one can easily show that subspaces of Banach spaces correspond to *sections of convex bodies*, and quotient spaces correspond to *projections of convex bodies.*

PROPOSITION 2.3 (Subspaces and quotient spaces). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a Banach space, and let $K = B_X$ denote is unit ball. Consider a subspace E of X. Then:*

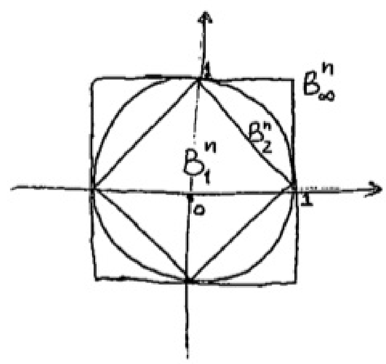

FIGURE 3. Unit balls of ℓ_p^n in dimension $n = 2$

1. The unit ball of E is the section $K \cap E$;

2. The unit ball of the quotient space X/E is isometrically equivalent to the orthogonal projection of K onto $E^\perp$, denoted $P_{E^\perp}(K)$.

The concept of dual space corresponds in convex geometry to the concept of polar set. Given a subset K of $\mathbb{R}^n$, the *polar set* K° is defined as

$$K^\circ = \{y \in \mathbb{R}^n : \langle x, y \rangle \leq 1 \text{ for all } x \in K\}.$$

Consider a Banach space $X = (\mathbb{R}^n, \|\cdot\|)$. By Riesz representation theorem, every linear functional $f \in X^*$ has the form $f(x) = \langle x, y \rangle = \sum_{i=1}^n x_i y_i$ for some $y \in \mathbb{R}^n$. Then the unit ball of X^* is

$$B_{X^*} = \{y \in \mathbb{R}^n : \langle x, y \rangle \leq 1 \quad \forall x \in B_X\} = (B_X)^\circ.$$

In words, the unit ball of the dual space X^* is the polar of the unit ball of X.

EXAMPLES 2.4. 1. Recall the duality relation $(\ell_1^n)^* = \ell_\infty^n$. This is a finite dimensional version of the duality considered in Example 1.2. The corresponding polarity relation is $(B_1^n)^\circ = B_\infty^n$, which says that the polar of a cross-polytope is a cube.

2. Similarly, $(B_2^n)^\circ = B_2^n$, i.e. the polar of a Euclidean ball is a Euclidean ball.

3. If $\mathcal{E}$ is an ellipsoid with semiaxes $a_1, \ldots, a_n > 0$ then Then $\mathcal{E}^\circ$ is an ellipsoid with semiaxes $1/a_1, \ldots, 1/a_n$.

4. If K is a polytope of k vertices and m faces, then K° is a polytope with m vertices and k faces.

Polarity goes well along with set theoretic operations.

PROPOSITION 2.5 (Properties of polar sets). *Let K and L be convex sets in $\mathbb{R}^n$. Then:*

1. $(K \cup L)^\circ = K^\circ \cap L^\circ$.

2. $(K \cap L)^\circ = \mathrm{conv}(K^\circ \cup L^\circ)$.
3. If $K \subset L$, *then* $L^\circ \subset K^\circ$.
4. (Bipolar theorem) If K *is closed then* $K^{\circ\circ} = K$.

Finally, we interpret the duality between subspaces and quotient spaces of Banach spaces described in Proposition 1.3 as a polarity relation between sections and projections of convex bodies.

PROPOSITION 2.6 (Duality between sections and projections). *Let* K *be a convex body in* $\mathbb{R}^n$, *and let* $E \subseteq \mathbb{R}^n$ *be a subspace. Let* P_E *denote the orthogonal projection in* $\mathbb{R}^n$ *onto* E. *Then:*
1. $(K \cap E)^\circ = P_E(K^\circ)$;
2. $(P_E(K))^\circ = K^\circ \cap E$.
In both parts, the operation of taking the polar set in the left hand side is considered in the subspace E.

PROOF. The inclusion $P_E(K^\circ) \subseteq (K \cap E)^\circ$ follows when we note that vectors $x \in K \cap E$ and $y \in K^\circ$ satisfy $\langle x, P_E y\rangle = \langle P_E x, y\rangle = \langle x, y\rangle \leq 1$. To prove the reverse inclusion, by Proposition 2.5 it suffices to show that $(P_E(K^\circ))^\circ \subseteq K \cap E$. Let $x \in (P_E(K^\circ))^\circ$; we have $\langle x, P_E y\rangle \leq 1$ for all $y \in K^\circ$. Since $x \in E$, we have $1 \geq \langle x, P_E y\rangle = \langle P_E x, y\rangle = \langle x, y\rangle$. It follows that $x \in K^{\circ\circ} = K$. This shows that $x \in K \cap E$. This completes the proof of Part 1.

Part 2 follows from part 1 and the bipolar theorem (Proposition 2.5). Indeed, we have $(K^\circ \cap E)^\circ = P_E(K^{\circ\circ}) = P_E(K)$. This completes the proof. ■

The following table summarizes the correspondence between functional analysis and convex geometry that we discussed in this section:

Spaces	*Unit balls*
normed space X, $\dim X = n$	symmetric convex body K in $\mathbb{R}^n$
subspace $E \subset X$	section $K \cap E$
quotient space X/E	projection $P_{E^\perp}(K)$
dual space X^*	polar body K°

3. A first glance at probabilistic methods: an approximate Caratheodory theorem

Geometric functional analysis heavily relies on probabilistic methods. In this section we will develop our first and simple probabilistic argument, which leads to what we might call an approximate Caratheodory's theorem.

We first recall the classical Caratheodory's theorem, which is a useful result in convex geometry.

THEOREM 3.1 (Caratheodory's theorem). *Consider a set* A *in* $\mathbb{R}^n$ *and a point* $x \in \mathrm{conv}(A)$. *There exists a subset* $A' \subseteq A$ *of cardinality* $|A'| \leq n+1$ *such that* $x \in \mathrm{conv}(A')$. *In other words, every point in the convex hull of* A *can be expressed as a convex combination of at most* $n+1$ *points from* A.

The bound $n+1$ can not be improved; it is clearly attained for a simplex A in $\mathbb{R}^n$ (a set of $n+1$ points in general position). However, if we only want to approximate x rather than exactly represent it as a convex combination, this is possible with much fewer points. Their number does not even depend on the dimension n, so the result holds in an arbitrary Hilbert space.

Let us fix an appropriate unit for approximation, which is the "radius" of A defined as

$$r(A) = \sup\{\|a\| :\ a \in A\}.$$

THEOREM 3.2 (Approximate Caratheodory's theorem). *Consider a bounded set A in a Hilbert space and a point $x \in \operatorname{conv}(A)$. Then, for every $N \in \mathbb{N}$, one can find points $x_1, \ldots, x_N \in A$ such that*

$$\Big\| x - \frac{1}{N} \sum_{i=1}^N x_i \Big\| \le \frac{r(A)}{\sqrt{N}}.$$

REMARKS. 1. Since $\frac{1}{N}\sum_{i=1}^N x_i$ is a convex combination of the points x_i, the conclusion of Theorem 3.2 implies that the point x almost lies in the convex hull of N points from A:

$$\operatorname{dist}\big(x, \operatorname{conv}(x_i)_{i=1}^N\big) \le \frac{r(A)}{\sqrt{N}}.$$

2. Theorem 3.2 has two advantages over the classical Caratheodory theorem. Firstly, Theorem 3.2 is dimension-free. Secondly, the convex combination to represent the arbitrary point is explicit: all the coefficients equal $1/N$. Note however that there may be repetitions among the points x_i.

PROOF. Let us fix a point $x \in \operatorname{conv}(A)$. We can express it as a convex combination of some vectors $z_1, \ldots, z_m \in A$ with some coefficients $\lambda_i \ge 0$, $\sum_{i=1}^m \lambda_i = 1$:

$$x = \sum_{i=1}^m \lambda_i z_i.$$

Consider a random vector Z which takes on the value z_i with probability λ_i for each $i = 1, \ldots, m$. Then

$$\mathbb{E} Z = \sum_{i=1}^m \lambda_i z_i = x.$$

Let $Z_1, Z_2, \ldots$ be independent copies of Z. By the the law of large numbers we know that the sum of independent random variables $\frac{1}{N}\sum_{j=1}^N Z_j$ converges to its mean x weakly (and even strongly) as N goes to infinity. The reason for this, as we may recall from the classical proof of the weak law of large numbers, is that the second moment of the difference is easily seen to converge to zero. Indeed, since the random variables $Z_i - x$ are independent and have mean zero, we obtain (see Exercise 1) that

$$\mathbb{E}\Big\|x - \frac{1}{N}\sum_{j=1}^{N} Z_j\Big\|^2 = \frac{1}{N^2}\mathbb{E}\Big\|\sum_{j=1}^{N}(Z_j - x)\Big\|^2 = \frac{1}{N^2}\sum_{j=1}^{N}\mathbb{E}\|Z_j - x\|^2.$$

Now, since the random variable X takes on values in A, we can easily bound each term of the sum (see Exercise 1):

$$\mathbb{E}\|Z_j - x\|^2 = \mathbb{E}\|Z - \mathbb{E}Z\|^2 = \mathbb{E}\|Z\|^2 - \|\mathbb{E}Z\|^2 \le \mathbb{E}\|Z\|^2 \le r(A)^2.$$

This shows that

$$\mathbb{E}\Big\|x - \frac{1}{N}\sum_{j=1}^{N} Z_j\Big\|^2 \le \frac{r(A)^2}{N}.$$

Therefore there is a realization of the random variables $Z_1, \dots, Z_N$ (i.e. a point in the probability space) such that

$$\Big\|x - \frac{1}{N}\sum_{j=1}^{N} Z_j\Big\| \le \frac{r(A)}{\sqrt{N}}.$$

Since by construction each Z_j takes on values in A, the proof is complete. ■

EXERCISE 1. 1. Let $Z_1, \dots, Z_N$ be independent mean zero random vectors in a Hilbert space. Show that

$$\mathbb{E}\Big\|\sum_{j=1}^{N} Z_j\Big\|^2 = \sum_{j=1}^{N}\mathbb{E}\|Z_j\|^2.$$

2. Let Z be a random vector in a Hilbert space. Show that

$$\mathbb{E}\|Z - \mathbb{E}Z\|^2 = \mathbb{E}\|Z\|^2 - \|\mathbb{E}Z\|^2.$$

This is a version of the identity $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ that holds for random variables X.

CHAPTER 2

Banach-Mazur distance

The concept of distance usually quantifies how far objects are from each other. Qualitatively, it can be used to distinguish the objects. For example, if two points in space have distance zero to each other, they are the same and thus can be identified with each other. The notion of distance is a quantitative version of identification; not only it can tell whether the objects are the same but it also measures how different they are.

How one identifies mathematical objects depends on their nature. For example, in topology, we identify spaces using homeomorphisms; in algebra, we identify groups using homomorphisms; in classical convex geometry, we may identify polytopes using rigid motions; in functional analysis, we identify Banach spaces using isomorphisms. The concept of Banach-Mazur distance which we shall study in this chapter is a quantitative version of the concept of isomorphism.

1. Definition and properties of Banach-Mazur distance

We will give two equivalent definitions of Banach-Mazur distance, one in the context of functional analysis, the other in convex geometry.

Definition 1.1. (Analytic): Let X, Y be two isomorphic normed spaces. The *Banach-Mazur distance* between X and Y is defined as

$$d(X,Y) = \inf \big\{ \|T\| \|T^{-1}\| : \ T : X \to Y \text{ is an isomorphism} \big\}.$$

(Geometric): Let K and L be symmetric convex bodies in $\mathbb{R}^n$. The *Banach-Mazur distance* between K and L is

$$d(K,L) = \inf \big\{ ab > 0 : \frac{1}{b} L \subseteq TK \subseteq aL \text{ for some linear operator } T \text{ on } \mathbb{R}^n \big\}.$$

Figure 1 illustrates the definition of Banach-Mazur distance between a hexagon and a circle.

The analytic and geometric notions of Banach-Mazur distance are connected through the correspondence between normed spaces and symmetric convex bodies described in Section 2. This way, the (analytic) Banach-Mazur distance between the spaces X and Y is the same as the (geometric) Banach-Mazur distance between their unit balls B_X and B_Y.

Remarks. 1. Recall that all finite-dimensional Banach spaces of the same dimension are isomorphic. Hence the notion of Banach-Mazur distance is well defined for them.

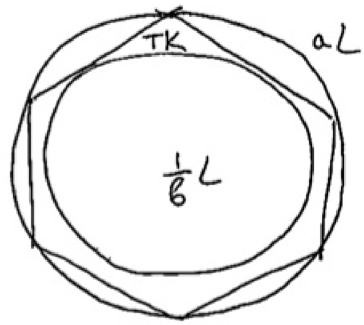

FIGURE 1. Banach-Mazur distance

2. Banach-Mazur distance is invariant under invertible linear transformations S: it is easily seen that $d(K, L) = d(SK, SL)$.

The following easy properties of Banach-Mazur distance follow from the corresponding properties of linear operators on normed spaces. We leave it to the reader to check them.

PROPOSITION 1.1 (Properties of Banach-Mazur distance). *Let X, Y, Z be n-dimensional normed spaces. Then:*

1. $d(X, Y) \geq 1$. Furthermore, $d(X, Y) = 1$ if and only if X and Y are isometric;
2. $d(X, Y) = d(Y, X)$;
3. $d(X, Y) \leq d(X, Z)d(Z, Y)$;
4. $d(X, Y) = d(X^*, Y^*)$.

The first three properties remind us of the conditions for a metric: non-negativity, symmetry, and triangle inequality. The only difference is that we have the above properties in multiplicative form. To get the additive form, we can simply take the logarithm:

COROLLARY 1.2 (Banach-Mazur Metric). $\log d(X, Y)$ *defines a metric on the set of all n-dimensional Banach spaces. Equivalently,* $\log d(K, L)$ *defines a metric on the set of all symmetric convex bodies in $\mathbb{R}^n$.*

Since $\log d(X, Y) = 0$ for isometric spaces X and Y, we feel that all isometric (but not isomorphic!) spaces should be identified with each other in geometric functional analysis. Geometrically this means, for example, that in $\mathbb{R}^n$ all ellipsoids are regarded the same, the cubes and paralleletopes are regarded the same, but an ellipsoid and a cube is not regarded the same.

One can check that the set of all n-dimensional Banach spaces equipped with the Banach-Mazur metric is a compact metric space. For this reason, this space is called the *Banach-Mazur compactum.* Very few properties of Banach-Mazur compactum are known, since it is usually hard to compute the Banach-Mazur distance between a given pair of normed spaces. However, when one of the spaces is ℓ_2^n, it is possible to give a tight upper bound $\sqrt{n}$

on the distance. We will now prove this remarkable result known as John's theorem.

2. John's Theorem

THEOREM 2.1 (John's theorem). *Let K be a symmetric convex body in $\mathbb{R}^n$, and let D be the ellipsoid of maximal volume contained in K. Then*

$$D \subseteq K \subseteq \sqrt{n}D.$$

Consequently, for arbitrary n-dimensional normed space one has

$$d(X, \ell_2^n) \leq \sqrt{n}.$$

PROOF. The second part of the theorem easily follows from the first part using the correspondence between normed spaces and convex bodies.

We are now set to prove the first part of the theorem. Since, as we know, Banach-Mazur distance is invariant under invertible linear transformations, we can assume that $D = B_2^n$ by applying an appropriate linear transformation to K and D.

We will proceed by contradiction. Assume that $K \not\subseteq \sqrt{n}D$. Then we can find $x_0 \in K$ such that

$$s := \|x_0\|_2 > \sqrt{n}. \tag{1}$$

By applying a rotation in $\mathbb{R}^n$ we can assume without loss of generality that the vector x_0 has the form $x_0 = (s, 0, \dots, 0)$. By symmetry and convexity of K, we have

$$\text{conv}\{\pm x_0, D\} \subseteq K.$$

We are going to find inside the convex body $\text{conv}\{\pm x_0, D\}$ an ellipsoid $\mathcal{E}$ of volume larger than D, which will contradict the maximality of D. We will construct $\mathcal{E}$ by significantly elongating the ball D in the direction of the s and slightly suppressing D in the orthogonal directions, see Figure 2.

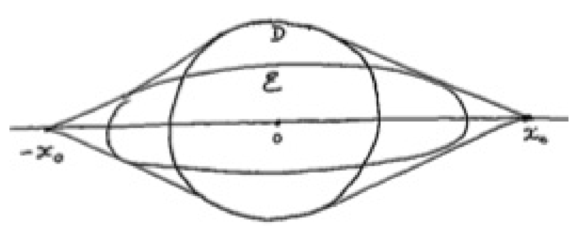

FIGURE 2. John's Theorem

To this end, let us consider the ellipsoid $\mathcal{E}$ centered at the origin, and whose semi-axes aligned in the coordinate directions have lengths $(a, b, b, \dots, b)$. Here $a, b > 0$ are parameters that we are going to determine.

Firstly, we would like to have $\mathrm{vol}(\mathcal{E}) > \mathrm{vol}(D)$. Since D is a unit ball, we have $\mathrm{vol}(\mathcal{E}) = ab^{n-1}\,\mathrm{vol}(D)$, which gives our first restriction

$$ab^{n-1} > 1.$$

Secondly, we would like to have $\mathcal{E} \subseteq \mathrm{conv}\{\pm x_0, D\}$, which we shall reformulate as a second constraint on a, b. Since both sides of this containment are bodies of revolution around the axis x_0, the problem becomes two-dimensional. At this point, the reader is encouraged to complete the proof independently using only planar geometry. For completeness, we proceed with the argument. Consider the quarter of any cross-section of these bodies orthogonal to x_0, see Figure 2.

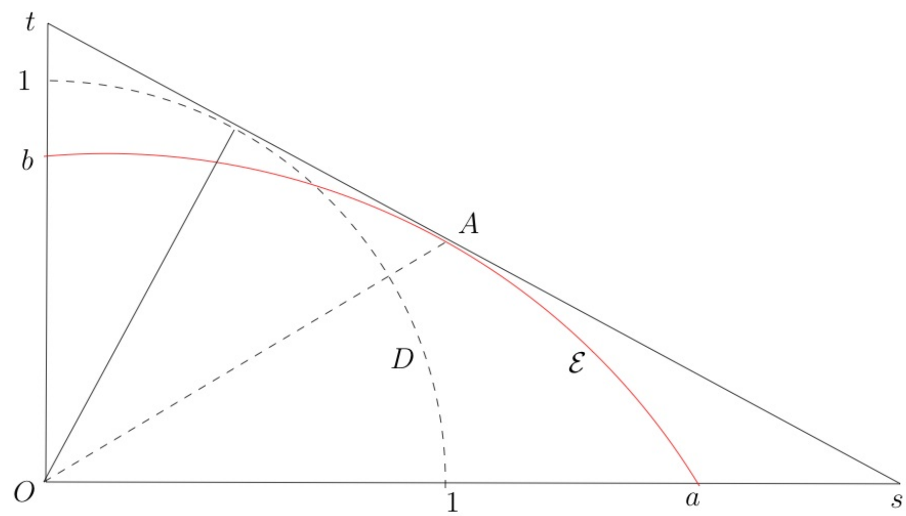

FIGURE 3. Proof of John's Theorem

We can clearly assume that both the ball D and the ellipsoid $\mathcal{E}$ are tangent to the boundary of $\mathrm{conv}\{\pm x_0, D\}$. Let t be the y-intercept of the line passing through x_0 (on x-axis with length s) and tangent to D. Similarity of triangles gives

$$\frac{s}{t} = \frac{\sqrt{s^2 - 1}}{1}. \tag{2}$$

Next, we are going to use the tangency of $\mathrm{conv}\{p, D\}$ and $\mathcal{E}$. To this end, we first shrink the picture along the x-axis by the factor b/a, see Figure 2. This transforms $\mathcal{E}$ into the Euclidean ball $\bar{\mathcal{E}}$ of radius b, and it transforms s into $\bar{s} = (b/a)\,s$. Similarity of triangles gives here

$$\frac{b}{\bar{s}} = \frac{t}{\sqrt{\bar{s}^2 + t^2}}.$$

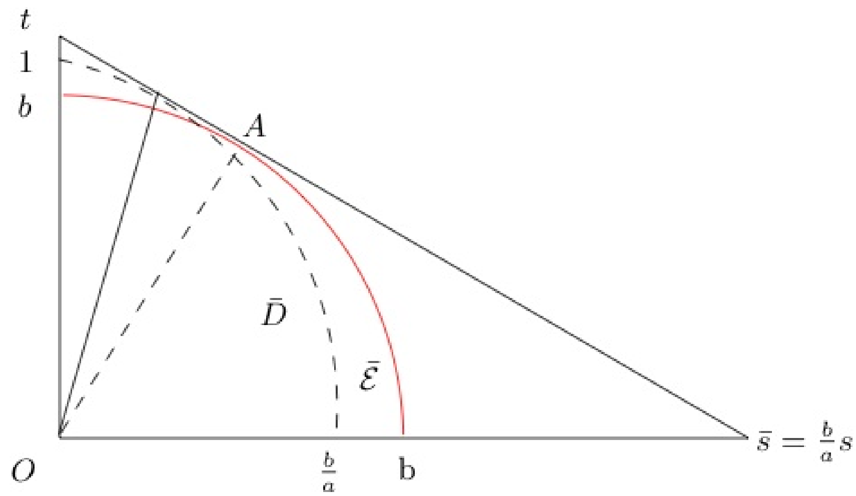

FIGURE 4. Proof of John's Theorem

This is equivalent to $s^2t^2 = b^2s^2 + a^2t^2$. Dividing both sides by t^2 and using (2) we conclude that $a^2 = s^2(1 - b^2) + b^2$. Let us choose $b^2 = 1 - \varepsilon$ for sufficiently small $\varepsilon > 0$; then $a^2 = 1 + (s^2 - 1)\varepsilon$. Since by our assumption $s > \sqrt{n}$, we conclude that (1) holds:

$$a^2b^{2(n-1)} > (1 + (n-1)\varepsilon)^2(1 - (n-1)\varepsilon) \geq 1.$$

John's Theorem is proved. ■

3. Distance between ℓ_p^n and ℓ_q^n

John's theorem is sharp. We will now show that a cube in n dimensions has the largest possible distance $\sqrt{n}$ to the ball.

PROPOSITION 3.1 (Sharpness of John's theorem). *For every $n = 1, 2, \dots$ we have*

$$d(\ell_\infty^n, \ell_2^n) = \sqrt{n}.$$

REMARKS. 1. By duality (Proposition 1.1), we also have $d(\ell_1^n, \ell_2^n) = \sqrt{n}$.

2. The geometric meaning of Proposition 3.1 is that among all ellipsoids, the round ball approximates the cube the best. While intuitively simple, this result is not entirely trivial, and it will be proved by a probabilistic argument.

PROOF. The upper bound $d(\ell_\infty^n, \ell_2^n) \leq \sqrt{n}$ follows easily from the norm comparison $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$, which in turn is a consequence of Hölder's inequality.

We are now going to prove the lower bound $d(\ell_\infty^n, \ell_2^n) \geq \sqrt{n}$. Note that in the geometric definition of Banach-Mazur distance we can require that $b = 1$ by rescaling. So it suffices to show that for every linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $B_2^n \subseteq T(B_\infty^n)$, we have $T(B_\infty^n) \not\subseteq sB_2^n$ for any

$s \leq \sqrt{n}$. The latter simply means that the radius of the parallelopiped $T(B_\infty^n)$ is large:

$$\max_{x \in B_\infty^n} \|Tx\|_2 > \sqrt{n}. \tag{3}$$

We are going to prove this by showing that a random vertex of the parallelopiped $T(B_\infty^n)$ has length larger than $\sqrt{n}$ with positive probability.

The vertices of the cube B_∞^n have the form $x = \sum_{i=1}^n \varepsilon_i e_i$, where $\varepsilon_i \in \{-1, 1\}$ and e_i denote the canonical basis vectors. Denote $f_i = Te_i$. Then the vertices of the parallelopiped $T(B_\infty^n)$ have the form

$$Tx = \sum_{i=1}^n \varepsilon_i T e_i = \sum_{i=1}^n \varepsilon_i f_i,$$

see Figure 3. Since $B_2^n \subseteq T(B_\infty^n)$, we must have $\|f_i\|_2 \geq 1$ for all i (why?).

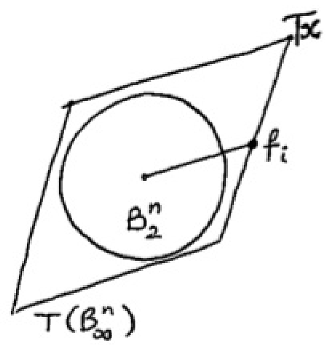

FIGURE 5. Banach-Mazur distance between a cube and a ball

Now we choose a random vertex Tx of the parallelepiped. Formally, choose the signs ε_i at random, so we let ε_i be independent symmetric $\{-1, 1\}$-valued random variables: $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$. By Exercise 1 and using the lower bound $\|f_i\|_2 \geq 1$, we can estimate the expectation

$$\mathbb{E}\|Tx\|_2^2 = \mathbb{E}\Big\| \sum_{i=1}^n \varepsilon_i f_i \Big\|_2^2 = \sum_{i=1}^n \|f_i\|_2^2 \geq n.$$

It follows that there exists choice of signs ε_i for which the random vertex Tx satisfies $\|Tx\|_2 \geq \sqrt{n}$. This establishes (3) and completes the proof. ■

As an application of Proposition 3.1, we now compute the distance between the spaces ℓ_p^n and ℓ_q^n in the case when the exponents p and q lie on the same side of 2.

COROLLARY 3.2 (Distance between ℓ_p^n and ℓ_q^n). *Let $1 \leq p \leq q \leq 2$ or $2 \leq p \leq q \leq \infty$. Then for every $n = 1, 2, \ldots$ we have*

$$d(\ell_p^n, \ell_q^n) = n^{1/p - 1/q}.$$

Proof. By duality (Proposition 1.1) we may assume that $2 \le p \le q \le \infty$. Then the upper bound $d(\ell_p^n, \ell_q^n) \le n^{1/p-1/q}$ follows from the norm comparison

$$\|x\|_q \le \|x\|_p \le n^{1/p-1/q}\|x\|_q$$

which in turn is a consequence of Hölder's inequality.

We will now deduce the lower bound $d(\ell_p^n, \ell_q^n) \ge n^{1/p-1/q}$ from the lower bound $d(\ell_2^n, \ell_\infty^n) \ge \sqrt{n}$ given by Proposition 3.1. One can do this simply by comparing the corresponding spaces in the two distance estimates. Indeed, the two special cases of the upper bound proved above are

$$d(\ell_2^n, \ell_p^n) \le n^{1/2-1/p}, \quad d(\ell_q^n, \ell_\infty^n) \le n^{1/q}.$$

Therefore, using the submultiplicativity of the Banach-Mazur distance (Proposition 1.1) we have

$$\begin{aligned}\sqrt{n} \le d(\ell_2^n, \ell_\infty^n) &\le d(\ell_2^n, \ell_p^n) \cdot d(\ell_p^n, \ell_q^n) \cdot d(\ell_q^n, \ell_\infty^n) \\ &\le n^{1/2-1/p} \cdot d(\ell_p^n, \ell_q^n) \cdot n^{1/q}.\end{aligned}$$

It follows that $d(\ell_p^n, \ell_q^n) \ge n^{1/p-1/q}$ as required. ■

It is a bit more difficult to compute the distance in the case when the exponents p and q lie on different sides of 2, that is for $1 \le p \le 2 < q \le \infty$. The identity map no longer realizes the distance between such spaces. For example, in the extreme case $p = 1$ and $q = \infty$ Hölder's inequality would give the upper bound on the distance $d(\ell_1^n, \ell_\infty^n) \le n$, which is far from being optimal. This distance is actually $\sqrt{n}$ up to an absolute constant; it is achieved by some rotation of the octahedron B_1^n inside B_∞^n that maps the vertices of the octahedron to some vertices of the cube. The general optimal bound for such range of p and q is

$$cn^\alpha \le d(\ell_p^n, \ell_q^n) \le Cn^\alpha, \qquad \text{where } \alpha = \max(1/p - 1/2,\ 1/2 - 1/q)$$

and where $C, c > 0$ are absolute constants, see [**?**, Proposition 37.6].

As an immediate application of John's theorem, we obtain a bound on the distance between any pair of n-dimensional normed spaces.

Corollary 3.3. *Let X and Y be n-dimensional normed spaces. Then*

$$d(X, Y) \le n.$$

Proof. The estimate follows from John's Theorem 2.1 and the submultiplicativity property of Banach-Mazur distance (Proposition 1.1): $d(X, Y) \le d(X, \ell_2^n) d(\ell_2^n, Y) \le \sqrt{n} \cdot \sqrt{n} = n$. ■

Corollary 3.3 is sharp up to a constant. This highly nontrivial result was proved by E. Gluskin in 1981. He constructed n-dimensional spaces X and Y for which $d(X, Y) \ge cn$. Gluskin's construction is randomized. He defines the unit ball of $X = (\mathbb{R}^n, \|\cdot\|)$ as the convex hull of $2n$ random points (and their opposites) on the unit Euclidean sphere of $\mathbb{R}^n$, taken independently and uniformly with respect to the Lebesgue measure. The second space

Y is constructed in the same way, independently from X. Heuristically, Banach-Mazur distance between such spaces X and Y should be large with high probability. This is because the vertices of B_X and B_Y are typically totally “mismatched”. The nontrivial part of Gluskin’s argument is to show that the vertices can not be sufficiently matched by any linear operator T on $\mathbb{R}^n$, which leads to the large Banach-Mazur distance. A detailed exposition of Gluskin’s construction, as well as many other results on Banach-Mazur distance, are in the monograph [**?**].

Part 1

Concentration of Measure and Euclidean Sections of Convex Bodies

The phenomenon of concentration of measure is the driving force of the early development of our subject – geometric functional analysis. It tells us that anti-intuitive phenomena occur in high dimensions. For example, we will see that the mass of a high dimensional ball is concentrated only on a thin band around any equator. This reflects the philosophy that metric and measure should be treated very differently: a set can have a large diameter but carries little mass. As we will see later on in this course, concentration of measure plays an important role in the cornerstone theorems such as the Dvoretzky's theorem and other theorems on sections or projections of convex bodies. In this lecture, we will study two cases of concentration of measure, each with a geometric form and a functional form:

- on the sphere
- in the Gauss space

4. Concentration of Measure

4.1. Concentration of measure on the sphere: geometric form.

Consider the unit sphere S^{n-1} with normalized Lebesgue measure $\sigma = \sigma_{n-1}$. Denote by E the equator and E_ϵ the ϵ-neighbourhood of the equator,

$$E_\epsilon = \{x \in S^{n-1} : d(x, E) \leq \epsilon\}.$$

First, we show that E_ϵ contains almost all the mass on the sphere when n is large.

Proposition 4.1 (Neighborhood of spherical caps).

$$\sigma(E_\epsilon) \geq 1 - 2e^{-n\epsilon^2/2}.$$

We will give an easy geometric proof for ϵ small.

Proof. Let C_ϵ denote the complement set of E_ϵ on the upper sphere. When ϵ is small, the "ice-cream" cone generated by C_ϵ is contained in a ball of radius $\sqrt{1-\epsilon^2}$ (See Figure 6):

$$\operatorname{cone}(C_\epsilon) \subset B\left(O', \sqrt{1-\epsilon^2}\right).$$

Then,

$$\begin{aligned}\sigma(C_\epsilon) &= \frac{\operatorname{vol}\left(\operatorname{cone}(C_\epsilon)\right)}{\operatorname{vol}(B_2^n)} \leq \frac{\operatorname{vol}\left(B(O', \sqrt{1-\epsilon^2})\right)}{\operatorname{vol}\left(B_2^n\right)} \\ &= \left(\sqrt{1-\epsilon^2}\right)^n \leq e^{-n\epsilon^2/2}.\end{aligned}$$

∎

Exercise 2. Make the treatment of ϵ rigorous: prove Proposition 4.1 for all $\epsilon > 0$.

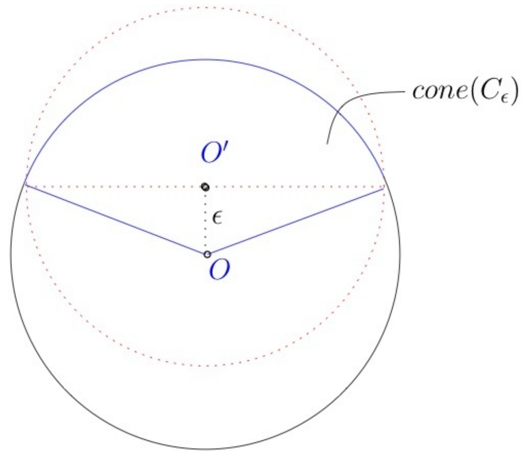

FIGURE 6. "ice-cream cone"

REMARKS. (1) The proposition implies that a band around the equator of width $\sim 1/\sqrt{n}$ has constant measure. Hence, as the dimension n gets larger and larger, the band that contains a fixed amount of mass becomes thinner and thinner.

(2) For any hemisphere A, we have $\sigma(A_\epsilon) \geq 1 - e^{-n\epsilon^2/2}$. In fact, this holds for any subset of measure $1/2$ (not necessarily the hemispheres). This follows from the classic isoperimetric inequality on the sphere.

THEOREM 4.2 (Isoperimetric inequality on the sphere: geometric form. [**?**], [**?**]). *Among all measurable sets $A \subset S^{n-1}$ with a given measure, spherical caps minimize the measure of the ϵ-neighborhood $\sigma(A_\epsilon)$.*

Consequently, we have the following theorem of concentration on the sphere.

THEOREM 4.3 (Concentration of measure on the sphere). *For an arbitrary measurable set $A \subset S^{n-1}$ with $\sigma(A) \geq \frac{1}{2}$,*

$$\sigma(A_\epsilon) \geq 1 - e^{-n\epsilon^2/2}.$$

REMARKS. (1) The number $1/2$ is not essential. We have the same concentration phenomenon (up to some constant adjustment) as long as the given measure is an absolute constant.

(2) The reason why Theorem 4.2 is called "isoperimetric inequality" is that, if we consider the measure of the boundary of A,

$$\sigma_{n-2}(\partial A) = \lim_{\epsilon \to 0} \frac{\sigma(A_\epsilon) - \sigma(A)}{\epsilon},$$

then the isoperimetric inequality implies that among all sets of a given volume, spherical caps have the minimal boundary measure.

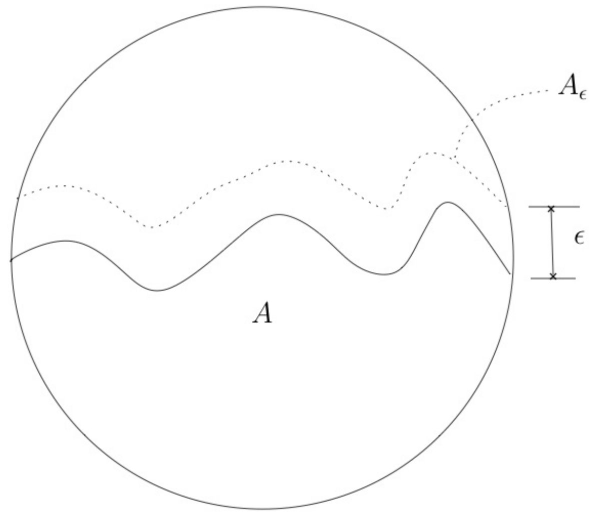

FIGURE 7. Concentration on the Sphere

(3) The sphere S^{n-1} locally looks like the Euclidean space $\mathbb{R}^n$, so we can get the isoperimetric inequality in $\mathbb{R}^n$ from that on the sphere.

COROLLARY 4.4. *(Isoperimetric inequality in $\mathbb{R}^n$) Among all sets of a given volume in $\mathbb{R}^n$, Euclidean balls minimize the surface area.*

Later in this course when we talk about geometric inequalities, we will come back to these isoperimetric inequalities and concentration of measure inequalities, and see alternative proofs.

4.2. Concentration of measure on the sphere: functional form. Concentration of measure is closely related to a special type of functions called Lipschitz functions. A general idea is that a Lipschitz function depending on many variables is almost a constant. A natural choice for constant is the median of the function. Let us recall some definitions first.

DEFINITION 4.1. Let X, Y be metric spaces. A map $f : X \to Y$ is called *L-Lipschitz* if

$$d(f(x), f(y)) \leq L \cdot d(x, y)$$

for all $x, y \in X$.

DEFINITION 4.2. Let X be a random variable. Then its *median* is a number M that satisfies

$$\Pr(X \leq M) \geq \frac{1}{2} \quad \text{and} \quad \Pr(X \geq M) \geq \frac{1}{2}.$$

Consider a 1-Lipschitz function defined on the sphere

$$f : S^{n-1} \to \mathbb{R}.$$

Let M be a median of f. That is,

$$\sigma\left\{x \in S^{n-1} : f(x) \geq M\right\} \geq \frac{1}{2} \quad \text{and} \quad \sigma\left\{x \in S^{n-1} \geq M\right\} \geq \frac{1}{2}.$$

Let $A = \{x \in S^{n-1} : f(x) \leq M\}$. Then $\sigma(A) \geq \frac{1}{2}$. By the concentration of measure on the sphere, we get

$$\sigma(A_\epsilon) \geq 1 - e^{-n\epsilon^2/2}.$$

It is easy to check that $f(x) \leq M+\epsilon$ for all $x \in A_\epsilon$ by the Lipschitz condition. Hence,

$$\sigma(f \leq M + \epsilon) \geq 1 - e^{n\epsilon^2/2}.$$

Similarly,

$$\sigma(f \leq M - \epsilon) \geq 1 - e^{n\epsilon^2/2}.$$

Take intersection of the two sets, and we get

$$\sigma(|f - M| \leq \epsilon) \geq 1 - 2e^{-n\epsilon^2/2}.$$

Thus, we have proved the following:

THEOREM 4.5. *(Concentration of measure on the sphere: functional form) Let $f : S^{n-1} \to \mathbb{R}$ be 1-Lipschitz with median M. Then,*

$$\sigma(|f - M| \leq \epsilon) \geq 1 - 2e^{-n\epsilon^2/2}.$$

REMARKS. Roughly speaking, this theorem says that a smooth function is almost constant on almost the entire sphere.

4.3. Concentration of measure in Gauss space: geometric form. We should note that in the formulation of the concentration of measure on the sphere, we did not quite use the geometry of the sphere. So it is not surprising that the same phenomenon happen in many other metric probability spaces. A classic example is the *Gauss space*.

DEFINITION 4.3. *Gauss space* is $\mathbb{R}^n$ equipped with the standard Gaussian measure $\gamma = \gamma_n$, with density

$$d\gamma_n = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2}.$$

Just as in the case of the sphere, we have an isoperimetric inequality in the Gauss space.

THEOREM 4.6 (Isoperimetric inequality in Gauss space). *Among all measurable sets $A \subset \mathbb{R}^n$ with a given Gaussian measure, half-spaces minimize $\gamma(A_\epsilon)$.*

Let H denote a half space. Then $H_\epsilon = \{x \in \mathbb{R}^n : x_1 \leq \epsilon\}$. A simple computation shows that

$$\gamma(H_\epsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\epsilon} e^{-x_1^2/2} dx_1 \geq 1 - e^{-\epsilon^2/2}.$$

Hence, we have the following concentration of measure theorem in the Gauss space.

THEOREM 4.7 (Concentration of measure in Gauss space: geometric form). *Let $A \subset \mathbb{R}^n$ be an arbitrary measurable set with $\gamma(A) \geq \frac{1}{2}$. Then $\gamma(A_t) \geq 1 - e^{-t^2/2}$.*

4.4. Concentration of measure in Gauss space: functional form. Just as in the case of the sphere, there is a functional form of concentration for Lipschitz functions in Gauss space. It can be established from the geometric form in the same way as above.

THEOREM 4.8 (Concentration of measure in Gauss space: functional form). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz with median M. Then for any $t > 0$,*

$$\gamma(|f - M| \leq t) \geq 1 - 2e^{-t^2/2}.$$

EXERCISE 3. Generalize the above concentration theorems (including functional forms) to L-Lipschitz functions.

A trouble here is that median is usually hard to compute. We would like to replace it by the mean for computational purposes. Fortunately, these two notions of average value are not far apart when the function has certain concentration properties.

LEMMA 4.9. *Let X be a random varible with mean $\mathbb{E}X$ and median M. Then $M \leq 2\mathbb{E}X$.*

PROOF. By Markov inequaluty,

$$\Pr(X > 2\mathbb{E}X) \leq \frac{1}{2} \leq \Pr(X \geq M).$$

If $M > 2\mathbb{E}X$, then

$$\Pr(X > M) \leq \Pr(X > 2\mathbb{E}X) < \frac{1}{2}.$$

This contradicts the median definition. ∎

LEMMA 4.10 (Mean $\approx$ Median under concentration hypothesis). *Let X be a random variable with mean $\mathbb{E}X$ and median M. Suppose that*

$$\Pr(|X - M| > t) \leq Ce^{-ct^2}$$

for any $t > 0$. Then

$$|M - \mathbb{E}X| \leq C_1.$$

Consequently,

$$\Pr(|X - \mathbb{E}X| > t) \leq C_2 e^{-c_2 t^2}$$

for all $t > 0$, where C_1, C_2, c_2 are constants that depend only on the absolute constants C, c.

PROOF.

$$
\begin{aligned}
|M - \mathbb{E}X| = |\mathbb{E}(M - X)| &\le \mathbb{E}|M - X| \quad \text{by Jensen's inequality} \\
&= \int_0^\infty \Pr(|X - M| > t)dt \le \int_0^\infty Ce^{-ct^2} dt = C_1.
\end{aligned}
$$

■

REMARKS. Any type of concentration with tail bound that makes the last integral converge will give the same "mean ≈ median" result up to some constant.

With this lemma, we have that, under appropriate concentration assumptions,

$$|X - \mathbb{E}X| \le |X - M| + C_1.$$

Assume $t > 2C_1$. This can always be done by choosing C_2 sufficiently large. Then,

$$
\begin{aligned}
\Pr(|X - \mathbb{E}X| > t) &\le \Pr(|X - M| > t - C_1) \\
&\le \Pr\left(|X - M| > \frac{t}{2}\right) \le Ce^{-c_2 t^2}.
\end{aligned}
$$

Hence, concentration of measure results on the sphere and in the Gauss space still hold when median is replaced by the mean.

EXERCISE 4. Show that the concentration results actually hold when the median is replaced by any p-th moment of f, $(\mathbb{E}f^p)^{1/p}$ for any $0 < p < \infty$ and $f \ge 0$.

As a point x in Gauss space is a random vector with independent standard Gaussian coordinates, we can derive a deviation inequality for certain functions of i.i.d Gaussian random variables. For a more detailed treatment on this topic, see [**?**].

COROLLARY 4.11 (Deviation Inequality). *Let $g_1, \cdots, g_n$ be independent standard Gaussian random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz. Then the random vector $X = f(g_1, \cdots, g_n)$ satisfies the deviation inequality*

$$\Pr(|X - \mathbb{E}X| > t) \le 2e^{-t^2/2} \quad \forall t > 0.$$

EXAMPLE 4.12. Consider the random variable

$$X = \sum_{i=1}^n a_i g_i,$$

where $\sum_{i=1}^n a_i^2 = 1$. It is easy to see that the linear functional

$$f(x) = \sum_{i=1}^n a_i x_i$$

is 1-Lipschitz. By Corollary 4.11, the normalized sum X satisfies the deviation inequality

$$\Pr(|X| > t) \leq 2e^{-t^2/2} \quad \forall t > 0.$$

This is a well-known fact, since normalized sum of standard Gaussians is standard Gaussian.

REMARKS. (1) The significance in this corollary is that the function f only needs to be "smooth": it does not have to be linear. Roughly speaking, this corollary says that a random variable that is a smooth function of many independent Gaussians strongly concentrates around its mean.

(2) Talagrand [**?**] proved a version of Corollary 4.11 for bounded random variales $|\xi_i| \leq c$.

(3) Other versions of Corollary 4.11 still remain open, even for subgaussians. A possible reason is that Lipschitz functions deal with Euclidean distances, and we lack a version of isoperimetry for subgaussian spaces. Another reason is that in Gauss space, there is usually heavy dependence on the rotation invariance of the Gaussian measure. However, for subgaussians, the group of symmetry is too small (usually only reflections, permutation of coordiantes). There may be hope if one can bypass the rotation invariance property to obtain concentration properties, for example, via log Sobolev ienqualities, etc.

5. Johnson-Lindenstrauss Flattening Lemma

Johnson-Lindenstrauss Flattening Lemma was motivated by data compression, or "dimension reduction." A lot of times, we need to store a large set of data in a complicated metric space. This would require a huge amount of space. For example, if we have n vectors in $\mathbb{R}^n$, we would need n^2 bits for exact storage. To reduce the complexity while still preserve the essential information of the original data set, we would love to embed the metric space into a simpler metric space almost isometrically. Here, "simpler" means smaller dimension. The question is, how far can we push the dimension down without losing the essential information? Johnson-Lindenstrauss flattening lemma tells us that to preserve the pairwise distance between the data points, the dimension can go down to the log level.

As a preliminary, we introduce a notion called "Lipschitz embedding" in metric spaces.

DEFINITION 5.1 (Lipschitz Embedding). Let X, Y be metric spaces. A map $T : X \to Y$ is called a *Lipschitz embedding* of X into Y if there exists $L > 0$ such that

$$\frac{1}{L} \cdot d(x, y) \leq d(Tx, Ty) \leq L \cdot d(x, y) \qquad \forall x, y \in X.$$

In other words, both T and T^{-1} (restricted on the image of T) are L-Lipschitz. Sometimes, such T is called *bi-Lipschitz.*

For example, all isometries are 1-embedding, for $d(Tx, Ty) = d(x, y)$ for all $x, y \in X$. An "*almost isometric embeddings*" means a $(1+\epsilon)$-embedding for small $\epsilon > 0$.

Now we are ready to state the flattening lemma.

THEOREM 5.1 (Johnson-Lindenstrauss Flattening Lemma). *Let X be an n-point set in a Hilbert space. For any $\epsilon > 0$, there exists a $(1+\epsilon)$-embedding of X into ℓ_2^k, where $k \geq C\epsilon^{-2}\log n$.*

REMARKS. Here, we consider the dimension reduction in ℓ_2. We want to preserve the pairwise Euclidean distances among all the data points. Note that we can always embed X in *span*(X) whose dimension $\leq n$. To store all pairwise distances in X, we would need $\binom{n}{2} \sim n^2$ numbers. Moreover, to store X itself, we would need $n \cdot n = n^2$ numbers (assuming each vector is n-dimensional). However, after embedding into ℓ_2^k, we only need to store $n \log n$ numbers while still preserving all pairwise distances within ϵ-error.

However, mo similar result holds in ℓ_1: one cannot do dimension reduction in ℓ_1. See [**?**].

We will use a linear operator G (in general, emeddings do not have to be linear), and our construction will be non-adaptive: it is independent of the geometry of X. This non-adaptivity is the radical idea of this proof. Note that in general, a good embedding should be adaptive. For example, if X consists of n co-linear points, then we do not want this line contained in $\ker(G)$: it would be perfect isometric embedding if this line is perpendicular to the kernel.

PROOF. Without loss of generality, assume that $X \subset \ell_2^n$. We need a map $G : \ell_2^n \to \ell_2^k$ which is a $(1+\epsilon)$-embedding when restricted to X. In this proof, we will use an $k \times n$ random Gaussian matrix G. Here is the plan of the proof. We first fix a pair $x, y \in X$ and check that

$$\|G(x-y)\|_2 \approx \|x-y\|_2.$$

Then, we take union bound over all n^2 pairwise distances.
Step 1: Fixed vector.
Let $x \in S^{n-1}$. Then

$$\begin{aligned}
\mathbb{E}\|Gx\|_2^2 &= \mathbb{E}\sum_{i=1}^{k} \langle \gamma_i, x\rangle^2 \qquad \text{where } \gamma_i\text{'s are the row vectors of } G \\
&= k\mathbb{E}\langle \gamma, x\rangle^2 \qquad \text{where } \gamma \text{ is a standard Gaussian vector in } \mathbb{R}^n, \\
&= k.
\end{aligned}$$

Now, consider the function defined on the Gauss space $\mathbb{R}^{kn}$:

$$f : \mathbb{R}^{kn} \to \mathbb{R}, G = (g_{ij})_{n\times n} \mapsto \|Gx\|_2.$$

We claim that f is 1-Lipschitz. In fact,

$$\begin{aligned}|f(G)-f(H)| &\leq |\|Gx\|_x - \|Hx\|_x| \leq \|(G-H)x\|_2 \\ &\leq \|G-H\| \leq \|G-H\|_{HS} = d(G,H),\end{aligned}$$

where $d(G,H)$ is the Euclidean distance of G and H when they are treated as elements of $\mathbb{R}^{kn}$.

Then, by Corollary 4.8, we have that, for every $x \in S^{n-1}$,

$$\Pr\left(\left|\|Gx\|_2 - \sqrt{\mathbb{E}\|Gx\|_2}\right| > t\right) \leq 2e^{-t^2/2} \qquad \forall t > 0,$$

where we have used the p-th moment replacement of the median in the concentration. That is,

$$\Pr\left(|\|Gx\|_2 - \sqrt{k}| > t\right) \leq 2e^{-t^2/2} \qquad \forall t > 0.$$

Let $t = \epsilon\sqrt{k}$ and replace G by $T := G/\sqrt{k}$. We get

$$\Pr\left(|\|Tx\|_2 - 1| > \epsilon\right) \leq 2e^{-\epsilon^2 k/2}.$$

Therefore, for arbitrary $x \neq 0$, we have

$$\Pr\left(|\|Tx\|_2 - \|x\|_2| > \epsilon\|x\|_2\right) \leq 2e^{-\epsilon^2 k/2}. \tag{4}$$

Step 2: Union bound.
Apply (4) to all $x := u - v$, where $u, v \in X$. We get

$$\Pr\left(\exists u, v \in X : |\|T(u-v)\|_2 - \|u-v\|_2| > \epsilon\|u-v\|_2\right) \leq n^2 \cdot 2e^{-\epsilon^2 k/2} \tag{5}$$

As long as the RHS is less than 1, that is, as long as $k \geq C\epsilon^{-2}\log n$, there exists a desired almost isometric embedding. ∎

REMARKS. (1) G can also be a random Bernoulli ± 1 matrix, as well as a sparse 0/1-matrix.
(2) Not much is know on derandomization of this argument.

Next, we introduce an alternative proof based on concentration of measure on the sphere. See section 15.2 in [**?**]. In this method, T will be an orthogonal projection in $\mathbb{R}^n$ onto a random k-dimensional subspace. First, let us be more explicit in what it means by a "random subspace."

DEFINITION 5.2 (Grassmannian). The *Grassmannian* $G_{n,k}$ is the collection of all k-dimensional subspaces of $\mathbb{R}^n$.

We will equip $G_{n,k}$ with a probability measure. For this purpose, we need the concept of *Haar measure* defined on a compact metric space.

Let (M, d) be a compact metric space, and let G be a group whose members act isometrically on M, i.e. $d(gx, gy) = d(x, y)$ for all $x, y \in M$ and $g \in G$. For example, $M = G_{n,k}$ equipped with Hausdorff distance ("angle" between the "orientations" of subspaces). Then $G = O(n) = \{\text{rotations and reflections}\}$, or one can take $G = SO(n) = \{\text{rotations}\}$.

THEOREM 5.2 (Haar measure).
1. *(Existence) There exists a Borel probability measure μ on M called* Haar measure *which is invariant under the action of G:*
$$\mu(gA) = \mu(A) \quad \forall g \in G, A \subseteq M.$$
2. *(Uniqueness) If the G-action is transitive, i.e. for any $x, y \in M$ there exists $g \in G$ such that $y = gx$, then the probability Haar measure is unique.*

EXAMPLE 5.3. $M = S^{n-1}$, $G = O(n)$. Then the Haar measure on M is the normalized Lebesgue measure.

EXAMPLE 5.4. $M = O(n) = G$. Then the Haar measure on M defines what we call "random rotations."

EXAMPLE 5.5. $M = G_{n,k}$, $G = O(n)$. The Haar measure defines what we call "random subspaces."

EXERCISE 5. Let U be a uniformly random rotation in $O(n)$ (i.e. pick a matrix from $O(n)$ uniformly at random).
1. Let $x \in S^{n-1}$ be fixed. Show that Ux is a a random vector uniformly distributed on S^{n-1}.
   Hint: Use the uniqueness of Haar measure on the sphere and show that Ux is rotation-invariant.
2. Let $E \in G_{n,k}$. Show that $U(E)$ is a random subspace uniformly distributed in $G_{n,k}$.

Now we are ready to show the alternative proof of the Johnson-Lindenstrauss lemma. Let T be a projection onto a random k-dimensional subspace. By the exercise above, this projection is equivalent to a random rotation of an orthogonal projection onto any fixed k-dimensional subspace. Hence, T can be realized as $U^* P_k U$, where U is uniformly random in $O(n)$, and P_k is the orthogonal projection onto $\mathbb{R}^k$. Moreover, for any fixed $x \in S^{n-1}$, Ux is uniformly random on S^{n-1}. Therefore,
$$\mathbb{E}\|Tx\|_2^2 = \mathbb{E}\|U^* P_k U x\|_2^2 = \mathbb{E}\|P_k z\|_2^2 = \mathbb{E}\langle z, P_k z\rangle = tr(P_k) = k.$$
This recovers the fixed vector step in the Gaussian proof above. The rest of the proof will be very similar to the Gaussian proof. That is, we apply the functional form of the concentration result on the sphere to the 1-Lipschitz function $f : S^{n-1} \to \mathbb{R}$ defined by $f(x) = \|Tx\|_2$, and then take union bound over all $x := u - v$. We leave the details to the reader.

6. Dvoretzky Theorem

6.1. Introduction. In the effort to understand the geometry of Banach spaces, it was noticed that not every Banach space has Hilbert subspaces (up to isomorphism). In particular, any subspace of ℓ_p has a further subspace isomorphic to ℓ_p, so ℓ_p has no Hilbert subspaces for $p \neq 2$, as every subspace of a Hilbert space is Hilbert.

Grothendieck's problem then asks, does every infinite dimensional Banach space have a finite dimensional subspace whose distance from a Hilbert space is at most of constant order, and whose dimension can be arbitrarily large?

In the early 1960's, Dvoretzky gave a positive answer to this question.

Theorem 6.1 (Dvoretzky's theorem). *Let X be an n-dimensional Banach space. Given any $\epsilon > 0$, there exists a subspace E of X of dimension $k = k(n, \epsilon) \to \infty$ as $n \to \infty$ such that $d(E, \ell_2^k) \leq 1 + \epsilon$.*

Remarks. We will see in the proof that k can be as large as $C(\epsilon) \log n$. Moreover, we will show that this is optimal for $X = \ell_\infty^n$.

One can interpret Dvoretzky's theorem geometrically as follows:

Theorem 6.2 (Geometric version of Dvoretzky's theorem). *Let K be a symmetric convex body in $\mathbb{R}^n$. Given any $\epsilon > 0$, there exists a section $K \cap E$ of K by a subspace E of $\mathbb{R}^n$ of dimension $k = k(n, \epsilon) \to \infty$ as $n \to \infty$ such that $\mathcal{E} \subseteq K \subseteq (1 + \epsilon)\mathcal{E}$ for some ellipsoid $\mathcal{E}$.*

Remarks. One can also insist that $\mathcal{E}$ is a round ball. See Lemma 14.4.1 in [**?**].

By using the duality between sections and projections (See the Preliminaries part of this course), we get a dual version of the Dvoretzky's theorem. This is obtained by interpreting the above theorem in the dual space.

Theorem 6.3 (Dual version of Dvoretzky's theorem). *Let X be an n-dimensional Banach space. Given any $\epsilon > 0$, there exists a quotient space X/F of dimension $k = k(n, \epsilon) \to \infty$ as $n \to \infty$ such that $d(X/F, \ell_2^k) \leq 1 + \epsilon$.*

The rest of this section will be organized as follows. We will first prove an intermediate result called the General Dvoretzky Theorem. It is a direct result of the concentration of measure on the sphere. This result gives the critical dimension of an almost Euclidean subspace in terms of the "average norm," the average value of the norm over the unit sphere. This is an intrinsic geometric attribute of the given normed space. Then, we will estimate the average norm in ℓ_p^n, and apply this General Dvoretzky Theorem to obtain some results regarding Euclidean subspaces of ℓ_p^n. In particular, we will show that an n-dimensional cube has Euclidean sections of dimension up to $c \log n$. To show that a general finite dimensional normed space admits an almost Euclidean subspace of large dimesion, all left to do is show that the average norm cannot be too small. To this end, we introduce the Dvoretzky-Rogers Lemma, which allows us to get a cube section of dimension up to $n/2$. Finally, by using the known result for ℓ_∞^n, we will obtain Dvoretzky Theorem.

6.2. General Dvoretzky Theorem. In our first attempt to prove Dvoretzky's theorem, we will introduce an extra parameter called the "average norm" in the expression of the subspace dimension. This becomes very natural when we apply the concentration of measure on the sphere.

Without loss of generality, assume $X = (\mathbb{R}^n, \|\cdot\|)$ and $\|x\| \le \|x\|_2 \ \forall x \in \mathbb{R}^n$. In other words, we assume that the unit Euclidean ball is contained in the unit ball of X. Then, consider the function $f : S^{n-1} \to \mathbb{R}$ defined by $f(x) = \|x\|$. Then f is 1-Lipschitz by our assumptions. By concentration of measure on the sphere, $f(x)$ concentrates closely around its average $\int_{S^{n-1}} \|x\| d\sigma$. This is what we refer to as the "*average norm.*" Let us give the statement of the General Dvoretzky Theorem now.

THEOREM 6.4 (General Dvoretzky's theorem). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space such that $\|x\| \le \|x\|_2$ for all $x \in \mathbb{R}^n$. Consider the average norm $M = \int_{S^{n-1}} \|x\| d\sigma(x)$. Then there exists a subspace E of dimension $k = c(\epsilon) n M^2$ such that*

$$(1-\epsilon) M \|x\|_2 \le \|x\| \le (1+\epsilon) M \|x\|_2$$

for all $x \in \mathbb{R}^n$. Consequently,

$$d(E, \ell_2^k) \le \frac{1+\epsilon}{1-\epsilon}.$$

The proof uses an important discretization argument. We will discretize the sphere using what we call "*ϵ-net,*" and use this net to approximate the entire sphere. The advantage of discretization by nets is that if we have good control over the size of the net, then we can use union bound to get from a fixed vector case to the general case. We have seen this line of thoughts before in the proof of Johnson-Lindenstrauss lemma.

Let us first pick up all the parts we will use later in the proof.

DEFINITION 6.1. Let $(\mathcal{M}, d)$ be a metric space, and let $\epsilon > 0$. A subset $\mathcal{N}$ of $\mathcal{M}$ is an ϵ-net if for all $x \in \mathcal{M}$, there exists some $y \in \mathcal{N}$ such that $d(x, y) \le \epsilon$.

REMARKS.
(1) Equivalently, $\mathcal{N}$ is an ϵ-net of $\mathscr{M}$ if $\mathcal{M}$ can be covered by ϵ-balls centered at $\mathcal{N}$.
(2) $\mathcal{M}$ is compact if and only if it has a finite ϵ-net.

Now we have the question: when $\mathcal{M}$ is compact, what is the minimal cardinality of an ϵ-net? In particular, here we are interested in the case where $\mathcal{M}$ is the unit ball B_2^n or the unit sphere S^{n-1}.

LEMMA 6.5 (cardinality of ϵ-net of the ball). *Let $\epsilon > 0$. Then there exists an ϵ-net $\mathcal{N}$ of B_2^n of cardinality*

$$|\mathcal{N}| \le \left(1 + \frac{2}{\epsilon}\right)^n.$$

Remarks. Just by comparing the volume, we can get a matching lower bound on the cardinality of an ϵ-net, which is also exponential in the dimension. Hence, the cardinality of a good ϵ-net of the unit ball is exponential in the dimension.

The proof of this lemma uses the correspondence between minimal covering and maximal packing.

Proof. Let $\mathcal{N}$ be a maximal ϵ-separated subset of B_2^n. That is,

$$d(x, y) \geq \epsilon \quad \forall x, y \in \mathscr{N}.$$

We claim that $\mathcal{N}$ is an ϵ-net of B_2^n. This can be easily justified by contradiction (otherwise, the maximality of the separated subset would be contradicted). Note that the balls with centers in $\mathcal{N}$ and radius $\epsilon/2$ are disjoint. Moreover, all these balls are contained $(1+\epsilon/2)B_2^n$. Hence, by volume comparison, we have

$$|\mathcal{N}| \cdot \operatorname{vol}\left(\frac{\epsilon}{2} B_2^n\right) \leq \operatorname{vol}\left(\left(1+\frac{\epsilon}{2}\right) B_2^n\right).$$

This gives the desired upper bound on $|\mathcal{N}|$. ■

Remarks. The same argument works for any unit ball K (instead of B_2^n) covered by its homothetic copy ϵK.

Exercise 6. Show that the same result holds for the unit sphere S^{n-1}.

Next, we want to approximate the operator norm using only the vectors in the net.

Lemma 6.6 (Computing operator norms on nets). *Let $T : X \to Y$ be a linear operator between normed spaces X and Y. Let $\mathcal{N}$ be a δ-net of the "sphere" $S_X := \{x \in X \ : \|x\| = 1\}$. Then,*

$$\|T\| \leq \frac{1}{1-\delta} \sup_{x \in \mathcal{N}} \|Tx\|.$$

Proof. For every $y \in S_X$, there exist $x \in \mathcal{N}$, $0 \leq \rho \leq \delta$ and $u \in S_X$ such that $y = x + \delta u$. Then,

$$\begin{aligned}
\|T\| &= \sup_{y \in S_X} \|Ty\| \\
&\leq \sup_{x \in \mathcal{N}} \|Tx\| + \rho \sup_{u \in S_X} \|Tu\| \\
&\leq \sup_{x \in \mathcal{N}} \|Tx\| + \delta \|T\|.
\end{aligned}$$

The lemma easily follows. ■

As a corollary of the lemma, we can also approximate $\inf_{y \in S_X} \|Ty\|$ using nets:

$$\inf_{y \in S_X} \|Ty\| \geq \inf_{x \in \mathcal{N}} \|Tx\| - \delta \|T\|. \tag{6}$$

Remarks. Since $\|T^{-1}\| = 1/\inf_{x \in S_X} \|Tx\|$, we see that nets are useful when we try to bound the condition number of T (which is defined as $\|T\| \cdot \|T^{-1}\|$).

The next corollary shows that if we get good control of vector norms on a net of the unit sphere, then we also have good control of vector norms over the entire unit sphere.

Corollary 6.7 (From nets to the sphere). *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, and let $\mathcal{N}$ be a δ-net of the unit sphere S^{n-1}. Suppose that for some $0 < \epsilon < 1$ and some $M > 0$, $(1-\epsilon)M \le \|x\| \le (1+\epsilon)M$ for all $x \in \mathcal{N}$. Then,*

$$(1-\epsilon-2\delta)M \le \|x\| \le \left(\frac{1+\epsilon}{1-\delta}\right) M$$

We leave it to the reader to work out the details of the proof. (Hint: The right hand side follows from Lemma 6.6, where we take T to be the identity operator from $(\mathbb{R}^n, \|\cdot\|_2)$ to $(\mathbb{R}^n, \|\cdot\|)$. For the left hand side, use (6).)

Exercise 7. We see that the upper bound is multiplicative while the lower bound is additive. Why can't we get a multiplicative lower bound such as $c(\delta) \cdot (1-\epsilon)M$?

Now we are ready to assemble all the parts in the proof of the General Dvoretzky Theorem.

Proof of General Dvoretzky Theorem. The proof will be completed in three steps.
Step 1: Control of a fixed vector norm by concentration.
The function $f : S^{n-1} \to \mathbb{R}$ defined by $f(x) = \|x\|$ is 1-Lipschitz by the assumption that $\|x\| \le \|x\|_2$. By the functional concentration of measure on S^{n-1}, we have

$$\sigma\left(|\|x\| - M| > t\right) \le 2\exp\left(-\frac{nt^2}{2}\right),$$

where $M = M_X = \int_{S^{n-1}} \|x\| d\sigma(x)$ is the average norm over the unit sphere. Choose $t = \epsilon M$ to get

$$\sigma\left(|\|x\| - M| > \epsilon M\right) \le 2\exp\left(-\frac{nM^2\epsilon^2}{2}\right).$$

In other words, the two sided estimate $(1-\epsilon)M \le \|x\| \le (1+\epsilon)M$ holds for a random point uniformly distributed over S^{n-1} with probability at least $1 - 2\exp\left(-nM^2\epsilon^2/2\right)$. *Step 2: Union bound over a δ-net.*
Let $\mathcal{N}$ be a δ-net of S^{k-1} with cardinality

$$|\mathcal{N}| \le \left(1 + \frac{2}{\delta}\right)^k \le \left(\frac{3}{\delta}\right)^k.$$

Let $U \in O(n)$ be uniformly random. Then for any fixed $x \in S^{n-1}$, Ux is uniformly random on the unit sphere S^{n-1}. In particular, Ux is uniformly random for any $x \in \mathcal{N}$. Step 1 gives that

$$(1-\epsilon)M \leq \|Ux\| \leq (1+\epsilon)M$$

with probability at least $1 - 2\exp\left(-nM^2\epsilon^2/2\right)$. Take union bound over $\mathcal{N}$, and we obtain

$$(1-\epsilon)M \leq \|Ux\| \leq (1+\epsilon)M \qquad \text{for all } x \in \mathcal{N}$$

with probability at least $1-|\mathcal{N}|\cdot 2\exp\left(-nM^2\epsilon^2/2\right)$. By the bound of $\mathcal{N}$, this probability $\geq 1-(3/\delta)^k \cdot 2\exp\left(-nM^2\epsilon^2/2\right)$. This is at least $1/2$ provided that $\delta = \epsilon$ and

$$k = \frac{c\epsilon^2}{\log(3/\epsilon)} nM^2 = c(\epsilon)nM^2. \tag{7}$$

Step 3: Approximation of the sphere by the δ-net.
Take the subspace E to be the image $U(\mathbb{R}^k)$ (so E is a random subspace). It is easy to check that $U(\mathcal{N})$ is an ϵ-net of $S^{n-1} \cap E$ (recall that $\delta = \epsilon$ is our choice from Step 2). By Cor 6.7,

$$(1-3\epsilon)M \leq \|z\| \leq \frac{1+\epsilon}{1-\epsilon}M \qquad \text{for all } z \in S^{n-1} \cap E.$$

Thus,

$$d(E, \ell_2^k) \leq \frac{1+\epsilon}{(1-\epsilon)(1-3\epsilon)} \leq 1 + 10\epsilon$$

for our choice of k as in (7). This finishes the proof of the general Dvoretzky theorem. ∎

Remarks. (1) Later we will show that

$$M \geq \sqrt{\frac{\log n}{n}}$$

for any symmetric convex body so that $k \sim \log n$.

(2) Notice that if $M \leq 1/\sqrt{n}$, then the theorem is meaningless because k would be a constant. Thanks to John's theorem, we need not worry about this. By John's theorem, we always have

$$\frac{1}{\sqrt{n}}\|x\|_2 \leq \|x\| \leq \|x\|_2$$

, so that $M \geq 1/\sqrt{n}$.

(3) Dvoretzky's theorem actually holds for "most" subspaces E in the sense that, if E is random from $G_{n,k}$, then the theorem holds with probability at least $1 - \exp(-c(\epsilon)k)$.

(4) The optimal dependence on ϵ is still open. The proof provides the dependence

$$c(\epsilon) \sim \frac{\epsilon^2}{\log\left(\frac{1}{\epsilon}\right)}.$$

(5) The "average norm" can actually be replaced by any p-th moment of the norm $\left(\int_{S^{n-1}} \|x\|^p d\sigma(x)\right)^{1/p}$.

(6) If we slightly relax the assumption to $\|x\| \le b\|x\|_2$ for all $x \in \mathbb{R}^n$, then Dvoretzk's theorem holds with

$$k \le c(\epsilon)n \left(\frac{M}{b}\right)^k.$$

(7) Given $\epsilon > 0$, the best k-value is called the *Dvoretzky dimension* of the space X, denoted by $k(X)$. It is the dimension of the largest Euclidean subspace that is $(1+\epsilon)$-close to a subspace of X.

Exercise 8. We can also obtain Dvoretzky's theorem by working with Gauss space instead of the unit sphere. Consider $\ell_X := \left(\int_{\mathbb{R}^n} \|x\|^2 d\gamma_n(x)\right)^{1/2} = \left(\mathbb{E}\|g\|^2\right)^{1/2}$, where g is the standard Gaussian vector in $\mathbb{R}^n$. Prove that

$$\ell_X = \sqrt{n}\left(\int_{S^{n-1}} \|x\|^2 d\sigma(x)\right)^{1/2} \sim \sqrt{n}M_X.$$

In fact, show that we can take $\ell_X = (\mathbb{E}\|g\|^p)^{1/p}$ for any $p > 1$, since $\left(\mathbb{E}\|g\|^2\right)^{1/2} \sim (\mathbb{E}\|g\|^p)^{1/p}$ for any $p > 1$.

Corollary 6.8 (General Dvoretzky Theorem: Gaussian formulation). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space such that $\|x\| \le b\|x\|_2$ for some $b > 0$ and for all $x \in \mathbb{R}^n$. Let $\ell_X = \left(\mathbb{E}\|g\|^2\right)^{1/2}$, where g is the standard Gaussian vector in $\mathbb{R}^n$. Then there exists a subspace E of dimension $k = c(\epsilon)\,(\ell_X/b)^2$ such that*

$$(1-\epsilon)\ell_X\|x\|_2 \le \|x\| \le (1+\epsilon)\ell_X\|x\|_2$$

for all $x \in \mathbb{R}^n$.

6.3. Euclidean subspaces of ℓ_p^n. As we mentioned earlier, to get from General Dvoretzky Theorem to Dvoretzky Theorem, we only need to estimate the average norm. In this section, we will compute the average norm in ℓ_p^n spaces and find the corresponding Dvoretzky dimensions. Note that we use the Gaussian formulation (Corollary 6.8) for computational purposes. We start from ℓ_1^n, a widely used structure in optimization problems.

Corollary 6.9 (Almost Euclidean subspace of ℓ_1^n). *There exists an almost Euclidean subspace of ℓ_1^n of dimension proportional to n. To be precise, there exists a subspace E of $\mathbb{R}^n$ of dimension $k \ge c(\epsilon)n$ such that*

$$(1-\epsilon)\|x\|_2 \le \frac{\|x\|_1}{\sqrt{n}} \le (1+\epsilon)\|x\|_2 \qquad \textit{for all } x \in E.$$

Proof. Note that

$$\|x\|_1 \le \sqrt{n}\|x\|_2 \quad \text{for all } x \in \mathbb{R}^n,$$

so $b = \sqrt{n}$ in Corollary 6.8. Moreover,

$$\ell_X = \mathbb{E}\|g\|_1 = \sum_{i=1}^{n} \mathbb{E}|g_i| = \sqrt{\frac{\pi}{2}}n.$$

Hence, the Dvoretzky dimension for ℓ_1^n is

$$k(\ell_1^n) = c(\epsilon)\left(\frac{\ell_X}{b}\right)^2 = c(\epsilon)n.$$

■

By duality, we immediately get the following:

Corollary 6.10 (Almost Euclidean quotient space of ℓ_∞^n). *There exists an almost Euclidean quotient space of ℓ_∞^n of dimension proportional to n. Equivalently, there exists an orthogonal projection of the unit cube $[-1,1]^n$ onto a k-dimensional subspace which is $(1+\epsilon)$-close to the Euclidean ball of radius $\sqrt{n}$, where $k = c(\epsilon)n$.*

Exercise 9. Show that $k(\ell_p^n) = cn$ for any $1 \le p \le 2$.

Next, we will find the Dvoretzky dimension of ℓ_q^n for $q \ge 2$.

Corollary 6.11 (Almost Euclidean subspaces of ℓ_q^n, $q \ge 2$). *There exists an almost Euclidean subspace of ℓ_q^n $(q \ge 2)$ of dimension $k = c(\epsilon)n^{2/q}$.*

Proof. Note that

$$\ell_X = \left(\mathbb{E}\|g\|_q^q\right)^{1/q} = \left(\sum_{i=1}^{n} \mathbb{E}|g_i|^q\right)^{1/q} = m_q n^{1/q},$$

where $m_q = (\mathbb{E}|g_1|^q)^{1/q}$ is the q^{th} moment of the standard Gaussian random variable, which is about $\sqrt{q}$. Hence, by Corollary 6.8,

$$k(X) = c(\epsilon)q \cdot n^{2/q}.$$

■

As a summary of Euclidean subspaces of ℓ_p^n spaces, we have the following.

Corollary 6.12 (Almost Euclidean Subspaces of ℓ_p^n). *Let X be a finite dimensional normed space. Denote by $k(X)$ the dimension of the largest Euclidean space that is $(1+\epsilon)$-isomorphic to a subspace of X. Then,*

- $k(\ell_p^n) \ge c(\epsilon)n, \qquad 1 \le p \le 2$
- $k(\ell_q^n) \ge c(\epsilon)q \cdot n^{2/q}, \qquad 2 \le q < \infty.$

Remarks. Both inequalities are sharp. See Section 5.4 in [**?**].

Next, we will obtain some results on the Dvoretzky dimension when we pair up a normed space with its dual space.

Proposition 6.13 (MM^* − *estimate*). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a finite dimensional normed space, and $X^* = (\mathbb{R}^n, \|\cdot\|_*)$ be the dual space of X. Let $M = M_X = \int_{S^{n-1}} \|x\| d\sigma(x)$ and $M^* = M_{X^*}$. Then $MM^* \ge 1$.*

PROOF. By Hölder's inequality,

$$\begin{aligned}(MM^*)^{1/2} &= \left(\int_{S^{n-1}} \|x\| d\sigma\right)^{1/2} \left(\int_{S^{n-1}} \|x\|_* d\sigma\right)^{1/2} \\ &\geq \int_{Sn-1} (\|x\|\|x\|_*)^{1/2}\, d\sigma \geq \int_{X^{n-1}} \langle x, x\rangle^{1/2} d\sigma = 1.\end{aligned}$$

Here, we have used the fact that $\|x\|_*$ can be recognized as the operator norm of the linear functional $f(z) = \langle z, x\rangle \quad \text{for all } z \in \mathbb{R}^n$. ■

COROLLARY 6.14. *Let $k(X)$ and $k(X^*)$ denote the Dvoretzky dimensions of X and its dual space X^*, respectively. Then*

$$k(X)k(X^*) \geq \frac{cn^2}{d(X, \ell_2^n)^2}.$$

PROOF. Assume that

$$a\|x\|_2 \leq \|x\| \leq b\|x\|_2$$

for some $a, b > 0$ such that

$$d(X, \ell_2^n) = \frac{b}{a}.$$

Then, by duality,

$$\frac{1}{b}\|x\|_2 \leq \|x\|_* \leq \frac{1}{a}\|x\|_2.$$

By General Dvoretzky Theorem,

$$k(X) \geq c(\epsilon)n \left(\frac{M}{b}\right)^2$$

and

$$k(X^*) \geq c(\epsilon)n \left(\frac{M^*}{1/a}\right)^2.$$

Hence,

$$k(X)k(X^*) \geq c(\epsilon)n^2 \left(\frac{MM^*}{b/a}\right)^2 \geq \frac{cn^2}{d(X, \ell_2^n)^2}$$

by Proposition 6.13. ■

REMARKS. By John's theorem, we always have $d(X, \ell_2^n) \leq \sqrt{n}$. Hence,

$$k(X)k(X^*) \geq cn$$

for some absolute constant $c > 0$. Consequently, for any finite dimensional normed space X, either

$$k(X) \geq c\sqrt{n}$$

or

$$k(X^*) \geq c\sqrt{n}.$$

Moreover, this conclusion is sharp: there exists an n-dimensional Banach space X such that both $k(X)$ and $k(X^*)$ are no less than $c\sqrt{n}$. See Section 3 in [**?**].

Finally, let us examine the Dvoretzky dimension of ℓ_∞^n. This will show the sharpness of Dvoretzky's Theorem. As a first step, we give a lower bound on $k(\ell_\infty^n)$.

PROPOSITION 6.15. *$k(\ell_\infty^n) \geq c \log n$, where $c > 0$ is an absolute constant.*

PROOF. Note that

$$\|x\|_\infty \leq \|x\|_2$$

for all $x \in \mathbb{R}$. So $b = 1$ in Corollary 6.8. We claim that

$$\ell_X = \mathbb{E} \max_{i \leq n} |g_i| \geq c\sqrt{\log n}. \tag{8}$$

To show that the expectation is large, it suffices to show the quantity is small only on a small portion of the space. Note that

$$\begin{aligned} \Pr(\max |g_i| \leq t) &= \Pr(|g_i| \leq t \text{ for all } i) \\ &= \Pr(|g_1| \leq t)^n \\ &\leq \left[1 - \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}\right]^n. \end{aligned}$$

By taking $t = c'\sqrt{\log n}$ we may obtain (8). Then, apply Corollary 6.8, and we get the desired lower bound on $k(\ell_\infty^n)$. ∎

Let us close this section with a summary for the Dvoretzky dimensions of ℓ_p^n for all $1 \leq p \leq \infty$.

- $k(\ell_p^n) \sim n$ if $1 \leq p \leq 2$;
- $k(\ell_q^n) \sim q \cdot n^{2/q}$ if $2 \leq q < \infty$;
- $k(\ell_\infty) \sim \log n$.

REMARKS. We have shown that $k(\ell_\infty^n) \geq c \log n$. We will establish the other direction $k(\ell_\infty^n) \leq c \log n$ in the next section.

6.4. Many faces of symmetric polytopes. In this section, we will show that for a convex symmetric polytope to well-approximate the Euclidean ball, it must have exponentially (in dimension) many faces. As a result, any symmetric convex polytope must have either $\sim e^{\sqrt{n}}$ vertices or $\sim e^{sqrtn}$ faces.

We start with two classic embedding theorems.

PROPOSITION 6.16 (Embedding into ℓ_∞).

(1) *Every separable Banach space X is isometric to a subspace of ℓ_∞ (possibly infinite dimensional).*

(2) *Ever Banach space X of dimension $k \sim c(\epsilon) \log n$ is $(1+\epsilon)$-isomorphic to a subspace of ℓ_∞^n.*

PROOF. (1) We need to find a linear isometric embedding $T : X \to \ell_\infty$.
Let $S_X = \{x \in X : \|x\| = 1\}$ denote the unit sphere of X. Since X is separable, there exists a countable, dense subset $(x_i)_{i=1}^\infty$ of S_X. By Hahn-Banach theorem, there exist $(u_i)_{i=1}^\infty$ in S_{X^*} such that $u_i(x_i) = 1$ for all i.
Define $Tx = \sum_{i=1}^\infty \langle x, u_i\rangle e_i$, where (e_i) is the standard basis vectors in ℓ_∞. Then, for every $x \in S_X$, one can easily check that

$$\|Tx\|_\infty = \sup_i |\langle x, u_i\rangle| = \sup_i |\langle x_k, u_i\rangle = 1,$$

where x_k approximates x. Hence, T is an isometry, as desired.

(2) The proof is similar to the first. Here, we take

$$\mathcal{N} = \{x_i, i = 1, \dots, n\}$$

to be an ϵ-net of S_X of cardinality

$$|\mathcal{N}| = n \leq \left(\frac{3}{\epsilon}\right)^k.$$

We leave the details to the reader.

■

REMARKS. It is well-known that every k-dimensional convex symmetric polytope with $2n$ faces is equivalent to a k-dimensional section of an n-dimensional cube. Hence, the second part of the proposition is equivalent to the following:

Every symmetric convex body of dimension $k = c(\epsilon) \log n$ is $(1 + \epsilon)$-isomorphic to a symmetric polytope with $2n$ faces.

A key ingredient in proving Dvoretzky's theorem is to reduce the general problem to the special case of the cube (i.e. ℓ_∞^n). We have obtained a lower bound on the Dvoretzky dimension in Proposition 6.15. Now we will derive a matching upper bound in order to conclude that $k(\ell_\infty^n) = c \log n$.

PROPOSITION 6.17. $k(\ell_\infty^n) \leq c \log n$.

PROOF. We want to find the right dimension k so that a k-section of the n-cube is $(1 + \epsilon)$-isomorphic to the k-dimensional Euclidean ball. Assume that P is a convex symmetric polytope in $\mathbb{R}^k$ such that

$$B_2^k \subseteq P \subseteq CB_2^k$$

for some constant $C > 0$. There exist $u_i \in \mathbb{R}^k$ (normal vectors to the faces) such that

$$P = \{x \in \mathbb{R}^k : |\langle x, u_i\rangle| \leq 1, i = 1, \cdots, n\}.$$

Since $B_2^k \subseteq P$, we have

$$\|u_i\|_2 \leq 1$$

for every i. On the other hand, as $P \subseteq CB_2^k$, there exists some i such that

$$|\langle x, u_i \rangle| \geq 1$$

for every $\|x\|_2 \geq C$. Equivalently, for every $x \in S^{k-1}$, there exists i such that

$$|\langle x, u_i \rangle| \geq \frac{1}{C}.$$

Without loss of generality, assume $\|u_i\|_2 = 1$ for all i. Consider the $1/C$-caps

$$C_i = \left\{ x \in S^{k-1} : \langle x, u_i \rangle \geq \frac{1}{C} \right\}.$$

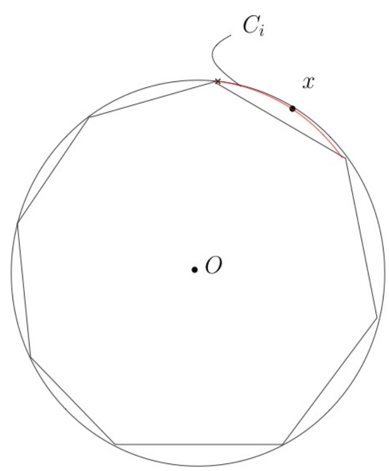

Figure 8. Many Faces of Symmetric Polytopes

Note that the unit sphere can be covered by n such caps, each corresponding to a face of the polytope P:

$$S^{k-1} \subseteq \bigcup_i C_i.$$

It is not hard to check that

$$\sigma(C_i) \leq e^{ck},$$

where c depends on C only. By comparing the volume, we will need $n \geq e^{ck}$ caps, and equivalently, this many faces. ■

Remarks. Consequently, the Euclidean ball B_2^k can be well approximated by a convex symmetric polytope with e^{ck} faces and no fewer. If we combine this fact and Corollary 6.14, we obtain the following result, which says that, given any n-dimensional convex symmetric polytope, either itself or its polar polytope has many faces.

COROLLARY 6.18 (Many faces of convex symmetric polytopes; [?]). *Let P be a convex symmetric polytope in $\mathbb{R}^n$. Suppose that P has v vertices and f faces (that is, $(n-1)$-dimensional faces). Then there is an abosolute constant $c > 0$ such that*

$$\log v \cdot \log f \geq cn.$$

PROOF. By the duality between vertices and faces, we know that the polar polytope P° has v faces and f vertices. By Proposition 6.17, we have that

$$k(P) = k(\ell_\infty^f) \leq c \log f,$$

and

$$k(P^\circ) = k(\ell_\infty^v) \leq c \log v.$$

Hence,

$$\log f \cdot \log v \geq ck(P)k(P^\circ) \geq cn.$$

■

6.5. Dvoretzky-Rogers Lemma. The last ingredient we need in the proof of Dvoretzky Theorem is Dvoretzky-Rogers Lemma. It helps to reduce the general problem to the case of ℓ_∞^n. It tries to "squeeze" an arbitrary Banach space between a cube from the outside and a Euclidean ball from the inside. In some sense, these two bodies are the two extremes in terms of Dvoretzky's dimension, ℓ_2^n being the best, and ℓ_∞^n the worst. Unfortunately, we cannot always "squeeze" the entire(meaning full-dimension) convex body like this. The factorization

$$\ell_2^n \to X \to \ell_\infty^n$$

with $T : \ell_2^n \to X$, $S : X \to \ell_\infty^n$ and $\|T\| \cdot \|S\| = O(1)$ does not always happen. See [?]. However, if we replace X with some subspace, we do have such factorization. Moreover, the subspace can have dimension up to $\frac{n}{2}$. This is done in Dvoretzky-Rogers Lemma.

LEMMA 6.19 (Dvoretzky-Rogers Lemma). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space. Assume that B_2^n is the ellipsoid of maximum volume in the unit ball B_X. Then there exists an orthonormal system $\{u_i : i = 1, \cdots, n/2\}$ in X such that*

$$\|u_i\| \geq \frac{1}{4}$$

for all $i = 1, \cdots, n/2$.

REMARKS. Note that $\|u_i\| \geq 1/4$ means that $u_i \in S^{n-1}$ cannot be too "deep" inside K: the deepest it can go is $1/4$ of the radius of K along this direction. In other words, the boundary point of K along the direction of u_i has Euclidean length at most 4. Therefore, the conclusion of Dvoretzky-Rogers lemma means that the section $K \cap \operatorname{span}\{u_i, i = 1, \ldots, n/2\}$ must be contained in the cube $[-4u_1, 4u_1] \times \cdots \times [-4u_{n/2}, 4u_{n/2}]$.

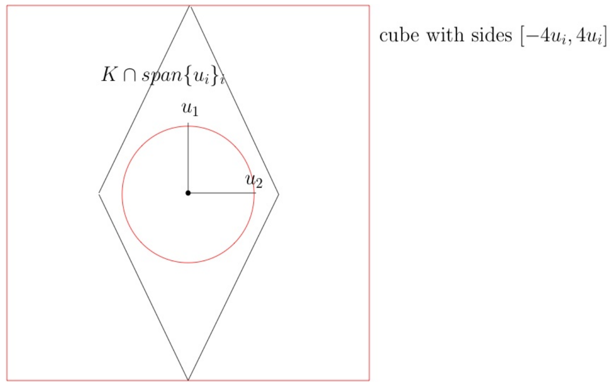

FIGURE 9. Dvoretzky-Rogers

PROOF. We will first construct the orthonormal system inductively. Let $u_1 \in S^{n-1}$ be any vector with the maximum norm in X. Clearly, $\|u_1\| = 1$, meaning that it is a contact point of B_X and B_2^n. Subsequently, for any $k \geq 2$, let $u_k \in S^{n-1} \cap \operatorname{span}\{u_i, i = 1, \dots, k-1\}^\perp$ be any vector with the maximum norm in X. Hence,

$$1 = \|u_1\| \geq \|u_2\| \geq \cdots \geq \|u_n\|.$$

The rest of the proof resembles the proof of John's theorem. The idea is that if $\|u_i\|$ is too small for some i, then $\|u_j\|$ would be too small for all $j > i$, which makes B_2^n too "deep" inside B_X in the directions of u_j for $j > i$. This would contradict the assumption that B_2^n is the ellipsoid of maximum volume: we could extend it in these directions and get an ellipsoid still contained in B_X but of larger volume.

Consider the following ellipsoid:

$$\mathcal{E} = \left\{ \sum_{j=1}^{n} a_j u_j : \frac{\sum_{j \leq n/2} a_j^2}{a^2} + \frac{\sum_{j > n/2} a_j^2}{b^2} \leq 1 \right\},$$

where $a = 1/2$ and $b = 1/(2\|u_{n/2}\|)$. We claim that $\mathcal{E} \subset B_X$. Then, since B_2^n has the maximal volume as an ellipsoid contained in B_X, we have

$$\operatorname{vol}(\mathcal{E}) = a^{n/2} b^{n/2} \operatorname{vol}(B_2^n) \leq \operatorname{vol}(B_2^n).$$

This gives our desired lower bound on $\|u_i\|$'s by plugging in the prescribed values for a and b.

It remains to prove the claim that $\mathcal{E} \subseteq B_X$. Let $\sum_{j=1}^n a_j u_j \in \mathcal{E}$. Then,

$$\sum_{j \leq n/2} a_j u_j \in B_2^n \subset B_X$$

and

$$\sum_{j > n/2} a_j u_j \in b B_2^n.$$

It follows that

$$\Big\| \sum_{j \le n/2} a_j u_j \Big\| \le a = \frac{1}{2}$$

and

$$\Big\| \sum_{j > n/2} a_j u_j \Big\| \le b \|u_{n/2}\|,$$

since $u_{n/2}$ was chosen to maximize the norm in X in $\operatorname{span}\{u_j, j > 2/n\} \cap S^{n-1}$. Then the claim follows directly from the triangle inequality. ■

6.6. Dvoretzky Theorem. Now we have all the ingredients in the proof of Dvoretzky Theorem. Just for recap: First, we proved the General Dvoretzky Theorem. It gives the Dvoretzky dimension of any n-dimensional Banach space in terms of an average norm:

$$k(X) = c(\epsilon) n (M_X/b)^2,$$

or equivalently,

$$k(X) = c(\epsilon)(\ell_X/b)^2,$$

where M_X is the average norm over S^{n-1}, and ℓ_X is the expected norm of a standard Gaussian vector. Then we showed that the Dvoretzky dimension of ℓ_∞^n is $c(\epsilon)\log n$. Later, we proved the Dvoretzky-Rogers Lemma, which allows us to embed an $n/2$-dimensional subspace of X into ℓ_∞^n. Hence, we can transfer our knowledge about ℓ_∞^n to X. In this section, we will complete this transfer and prove Dvoretzky Theorem. As we did before, it only remains to come up with the right estimate on ℓ_X.

Lemma 6.20 (ℓ_X from Dvor-Rogers lemma). *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space. Assume that B_2^n is the ellipsoid of maximum volume in the unit ball B_X. Then*

$$\ell_X \ge c\sqrt{\log n},$$

where $\ell_X = \mathbb{E}\|g\|$.

Proof. Let $E = \operatorname{span}\{u_i : i \le n/2\}$, where u_j's are from Dvoretzky-Rogers Lemma. Then, since g_i's are identically distributed as $\epsilon_i g_i$, where ϵ_i's are independent ± 1 equal probability Bernoulli random variables, we have

$$\begin{aligned}
\ell_E &= \mathbb{E}\|\sum_{n=1}^{n/2} \epsilon_i g_i u_i\| \\
&\ge \mathbb{E} \max_{1 \le i \le n/2} \|g_i u_i\| && (9)\\
&\ge \frac{1}{4}\mathbb{E} \max_{1 \le i \le n/2} |g_i| && (10)\\
&\ge c\sqrt{\log n}.
\end{aligned}$$

Inequality (9) comes from conditioning on the g_i's and applications of Fubini and Jensen's inequality. Note that

$$\begin{aligned}\mathbb{E}_{1,\dots,n/2}\Big\|\sum_{i=1}^{n/2}\epsilon_i v_i\Big\| &= \mathbb{E}_1 E_{2,\dots,n/2}\Big\|\sum_{i=1}^{n/2}\epsilon_i v_i\Big\| \\ &\geq \mathbb{E}_1\Big\|\epsilon_1 v_1 + \mathbb{E}_{2,\dots,n/2}\sum_{i=2}^{n/2}\epsilon_i v_i\Big\| \\ &= \mathbb{E}_1\|\epsilon_1 v_1\| = \|v_1\|.\end{aligned}$$

This works for any v_i, so we get (9). The second inequality (10) is a result of Dvoretzky-Rogers lemma. Hence, the proof is complete. ∎

Now, Dvoretzky's theorem is immediate. The theorem asserts that the Dvoretzky dimension of an arbitrary n-dimensional normed space X is at least $c(\epsilon)\log n$.

Proof of Theorem 6.2. By Lemma 6.20,

$$\ell_X \geq c\sqrt{\log n}.$$

Then, General Dvoretzky's theorem gives

$$k(X) \geq c(\epsilon)\ell_X^2 \geq c(\epsilon)\log n.$$

∎

7. Volume Ratio Theorem

In Dvoretzky's theorem, we see that every n-dimensional Banach space has an almost Euclidean subspace of logarithmic dimension. There we sacrifice on the dimension in order to get the subspace $1+\epsilon$-close to Euclidean for any prescribed value of ϵ. In this section, we will explore the other direction of the same picture: what if we would like to sacrifice on the distance to Euclidean in order to make the subspace almost full-dimension? This was first examined by Kashin in 1977, then simplified and generalized by Szarek in 1979. We will present an exposition inspired by [**?**] based on covering numbers.

7.1. Covering numbers.

Definition 7.1 (Covering number). Let K, D be two sets in $\mathbb{R}^n$. The *covering number* $N(K,D)$ is the minimum number of traslates of D needed to cover K.

Note that covering number is closely related to the notion of an ϵ-net.

Example 7.1. $N(K,\epsilon B_2^n)$ = minimum cardinality of an ϵ-net of K in Euclidean metric.

PROPOSITION 7.2 (Covering number estimate). *Let $K \subset \mathbb{R}^n$, and let $D \subset \mathbb{R}^n$ be a convex symmetric set. Then*

$$\frac{\mathrm{vol}(K)}{\mathrm{vol}(D)} \leq N(K, D) \leq \frac{\mathrm{vol}(K + \frac{1}{2}D)}{\mathrm{vol}(\frac{1}{2}D)}.$$

PROOF. The lower bound follows directly from a volume comparison. K can be covered by $N(K, D)$ translates of D, so

$$\mathrm{vol}(K) \leq N(K, D)\,\mathrm{vol}(D).$$

The upper bound can be obtained in a similar way to the estimate on the cardinality of ϵ-nets. See the proof of Lemma 6.5. ■

As an immediate consequence, we can get an estimate on the covering number of a convex body covered by its homothetic copies.

COROLLARY 7.3. *For any symmetric convex set $K \subset \mathbb{R}^n$ and every $\epsilon > 0$,*

$$\left(\frac{1}{\epsilon}\right)^n \leq N(K, \epsilon K) \leq \left(2 + \frac{1}{\epsilon}\right)^n.$$

7.2. Volume ratio theorem via entropy: local version. As we mentioned in the introduction part of this section, we want the dimension of the subspace arbitrarily close to full dimension, and find out how much we need to sacrifice on the "Euclideanness" of that subspace. It turns out that our sacrifice depends on a geometric quantity of the space called the "volume ratio."

THEOREM 7.4 (Volume ratio theorem (VRT for short)). *Let K be a convex symmetric body in $\mathbb{R}^n$. Suppose that $B_2^n \subset K$. Define the* volume ratio *by*

$$v(K) := \left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(B_2^n)}\right)^{\frac{1}{n}}.$$

Then, for every $\delta \in (0, 1)$, there exists a subspace E of dimension $(1 - \delta)n$ such that

$$\mathrm{diam}(K \cap E) \leq (cv(K))^{\frac{1}{\delta}}.$$

Equivalently,

$$B_2^n \cap E \subseteq K \cap E \subseteq (cv(K))^{\frac{1}{\delta}}\, B_2^n \cap E.$$

REMARKS.
(1) It is easy to see that the volume ratio is the same for all homothetic copies of a convex body.
(2) It implies that the section $K \cap E$ is within distance $(cv(K))^{\frac{1}{\delta}}$ from Euclidean.
(3) Note that the conclusion only depends on the volume ratio $v(K)$ and δ, not on the dimension n.
(4) The proof will show that VRT actually holds for a random subspace $E \in G_{n,(1-\delta)n}$ with high probability (exponential in n).

(5) It is still open whether we can achieve a polynomial dependence on δ if K is first put in a nice position. For example, is it true if K is in John's position, meaning that the ellipsoid of maximum volume is B_2^n? In general, this is not true: one can check that e will get exponential dependence when K is an ellipsoid with a degenerated direction (i.e. K looks like a "sausage").

The following corollary is simply a re-interpretation of VRT.

Corollary 7.5 (Almost Euclidean subspaces of finite-dimensional normed spaces). *Let X be an n-dimensional Banach space. Consider the volume ratio*

$$v(X) = \inf \left\{ \frac{\operatorname{vol}(B_X)}{\operatorname{vol}(\mathcal{E})} : \mathcal{E} \subset B_X \text{ is an ellipsoid} \right\}.$$

Then, for every $\delta \in (0,1)$, there exists a subspace $E \subset X$ of dimension $(1-\delta)n$ such that $d(E, \ell_2^{(1-\delta)n}) \leq (cv(X))^{\frac{1}{\delta}}$.

As we announced earlier, we will derive VRT from the view of covering numbers (sometimes referred to as "entropy"). The following theorem on entropy is the main step toward VRT.

Theorem 7.6 (Entropy). *Let K be a symmetric convex body in $\mathbb{R}^n$. Let v be the number such that*

$$v^n = N(K, B_2^n).$$

Then for any $\delta \in (0,1)$, there exists a subspace E of dimension $(1-\delta)n$ such that

$$\operatorname{diam}(K \cap E) \leq (cv)^{\frac{1}{\delta}}.$$

Remarks. The proof will show that, in fact, a random subspace E satisfies this with high probability.

First, let us see how this theorem immediately implies V.R.T.

Proof from entropy to VRT. By Theorem 7.6, it suffices to show $v \leq Cv(K)$. That is,

$$N(K, B_2^n) \leq \frac{\operatorname{vol}(CK)}{\operatorname{vol}(B_2^n)}.$$

This follows easily from Proposition 7.2and the fact that $B_2^n \subset K$.

$$\begin{aligned} N(K, B_2^n) &\leq \frac{\operatorname{vol}(K + \frac{1}{2}B_2^n)}{\operatorname{vol}(\frac{1}{2}B_2^n)} \\ &\leq \frac{\operatorname{vol}(\frac{3}{2}K)}{\operatorname{vol}(\frac{1}{2}B_2^n)} \\ &= \frac{3^n \operatorname{vol}(K)}{\operatorname{vol}(B_2^n)} \\ &= (3v(K))^n. \end{aligned}$$

In other words, we have $v \leq 3v(K)$. ■

REMARKS. The other side of Proposition 7.2 gives that $v(K) \leq v$. So we actually have

$$v(K) \leq v \leq 3v(K).$$

Hence, the two notions v and $v(K)$ are equivalent.

Before we go on to the proof of the entropy theorem, let us first picture the ideas. From the volume distribution in convex bodies (which decays exponentially away from the center), we can safely think of a high dimensional convex body as a ball with a few "tentacles" – points in the convex body that have large Euclidean norm. See Figure 10. The conclusion of the entropy theorem guarantees that there is a subspace that does not intersect any of these tentacles. In fact, by the remark following Theorem 7.6, most subspaces avoid these tentacles.

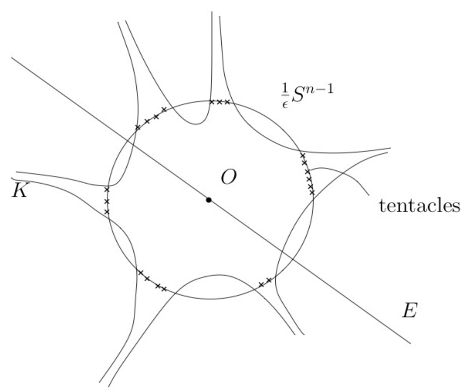

FIGURE 10. Central Ball and Tentacles

Here is the plan of the proof: First, we cover the tantacles by "not too many" unit balls, using the assumption that $N(K, B_2^n) = v^n$; then we show that a random subspace E is disjoint from any given covering ball with high probability; lastly, we use union bound over all covering balls to get that a random subspace E is disjoint from all the covering balls with positive (in fact, high) probability.

We have everything ready to realize this plan except the second step. Informally, it says that a random subspace is, with high probability, not too close to a fixed point.

PROPOSITION 7.7 (Distance of a Random Subspace to a Fixed Point). *Let $E \in G_{n,n-k}$ be a random subspace* (codim $E = k$)*. Let $x \in S^{k-1}$ be arbitrary (but fixed). Then,*

(1) $\left(\mathbb{E}\operatorname{dist}(x,E)^2\right)^{\frac{1}{2}} = \sqrt{\frac{k}{n}}$;

(2) $\Pr\left(\operatorname{dist}(x,E) \leq \epsilon\sqrt{\frac{k}{n}}\right) \leq (c\epsilon)^k$ *for any* $\epsilon > 0$.

We will present the proof of the first part, and leave the second as an exercise.

Proof. We will use an equivalent model: by the uniqueness of Haar measure, consider E as a fixed subspace, say $E = \mathbb{R}^{n-k}$ and $x \in S^{n-1}$ a random uniform vector.

(1) Let $x = (x_i)_{i=1}^n \in S^{n-1}$. Note that all coordinates of x are identically distributed (but not independent). Then,

$$1 = \|x\|_2^2 = \mathbb{E}\|x\|_2^2 = n\mathbb{E}x_1^2,$$

so that $\mathbb{E}x_1^2 = \frac{1}{n}$. Hence,

$$\mathbb{E}\,\mathrm{dist}(x,E) = \mathbb{E}\|P_k x\|_x = \mathbb{E}\sum_{i=1}^{k} x_j^2 = k\mathbb{E}x_1^2 = \frac{k}{n},$$

where P_k is the orthogonal projection onto the k-dimensional subspace $E^{\perp}$.

(2) Consider the k-dimensional "band" around the $(n-k)$-dimensional equator S^{n-k} on S^{n-1}:

$$\Pr\Big(\mathrm{dist}(x,E) \le \epsilon\sqrt{\frac{k}{n}}\Big) = \sigma\Big\{\mathrm{dist}(x, \mathbb{R}^{n-k} < \epsilon\sqrt{\frac{k}{n}}\Big\}.$$

■

Exercise 10. Show that $\sigma(A) \sim (c\epsilon)^k$, where

$$A = \{x \in S^{n-1} : \mathrm{dist}(x,E) \le \epsilon\sqrt{\frac{k}{n}}\}.$$

(Hint: Use polar coordinate.) For sharp estimate of $\sigma(A)$, see [**?**].

Now, we are ready to realize our plan for the proof of VRT.

Proof of VRT. Let $\epsilon > 0$ be fixed. Later, we will choose ϵ such that

$$\frac{1}{\epsilon} = (cv)^{\frac{1}{\delta}}.$$

We want to show

$$K \cap E \subset \frac{1}{\epsilon}B_2^n.$$

Equivalently,

$$K \cap \frac{1}{\epsilon}S^{n-1} \cap E = \emptyset.$$

We will prove this in three steps.

Step 1: Discretize tentacles.

Let $\mathcal{N}$ be a 1-net of $K \cap \frac{1}{\epsilon}S^{n-1}$ in the Euclidean metric. Then,

$$|\mathcal{N}| \le N(K, B_2^n) = v^n.$$

Step 2: Fixed vector.
Given any $x \in K \cap \frac{1}{\epsilon} S^{n-1}$, we show that a random subspace E is far from x with high probability. By Prop 7.7

$$\begin{aligned}\Pr\Big((x + B_2^n) \cap E \neq \emptyset\Big) = \Pr\Big(\operatorname{dist}(x, E) \leq 1\Big) = \Pr\left(\operatorname{dist}\left(\frac{x}{\|x\|_2}, E\right) < \epsilon\right) \\ \leq \left(C\epsilon\sqrt{\frac{n}{k}}\right)^k = \left(\frac{C\epsilon}{\sqrt{\delta}}\right)^{\delta n}.\end{aligned}$$

Step 3: Union bound.

$$\begin{aligned}\Pr\Big(\forall x \in \mathcal{N}, (x + B_2^n) \cap E = \emptyset\Big) &\geq 1 - |\mathcal{N}| \left(\frac{C\epsilon}{\sqrt{\delta}}\right)^{\delta n} \\ &\geq 1 - \left[v\left(\frac{C\epsilon}{\sqrt{\delta}}\right)^{\delta}\right]^n \geq 1 - e^{-n},\end{aligned}$$

provided that $v\left(C\epsilon/\sqrt{\delta}\right)^{\delta} \leq 1/\delta$. Choose $\epsilon = 1/(Cv)^{1/\delta}$, and this completes the proof. ∎

EXAMPLE 7.8 (Euclidean sections of ℓ_1^n). First, we show that ℓ_1^n has uniformly bounded volume ratio. To this end, we need to compute $\operatorname{vol}(B_1^n)$ and $\operatorname{vol}(B_2^n)$.

$$\operatorname{vol}(B_1^n) = 2^n, \quad \operatorname{vol}(T_n) = \frac{2^n}{n!},$$

where T_n denotes the standard n-dimensional simplex $\operatorname{conv}\{0, e_1, \dots, e_n\}$. Next,

$$\operatorname{vol}(B_2^n) \geq \operatorname{vol}\left(\frac{1}{\sqrt{n}} B_\infty^n\right) = \left(\frac{2}{\sqrt{n}}\right)^n.$$

Let $K = \sqrt{n} B_1^n$ so that B_2^n is the ellipsoid of maximum volume. Then

$$\operatorname{vol}(K) = \frac{2^n n^{n/2}}{n!} \sim \left(\frac{c}{\sqrt{n}}\right)^n,$$

and thus

$$v(K) = \left(\frac{\operatorname{vol}(K)}{\operatorname{vol}(B_2^n)})\right)^{1/n} \leq \text{constant}.$$

Hence, by VRT, for any $0 < \delta < 1$, there exists a subspace E of ℓ_1^n whose dimension is $(1 - \delta)n$ such that

$$d\Big(E, \ell_2^{(1-\delta)n}\Big) \leq \text{constant}.$$

This was observed by Kashin in 1977.

Corollary 7.9 (Kashin). *For every $\delta \in (0,1)$, a random subspace $E \subset \mathbb{R}^n$ of codimension δn satisfies the following with probability at least $1-e^{-n}$:*

$$B_2^n \cap E \subset \sqrt{n} B_1^n \cap E \subset C^{\frac{1}{\delta}} B_2^n \cap E.$$

Remarks. (1) Equivalently,

$$C^{-\frac{1}{\delta}} \|x\|_2 \leq \frac{\|x\|_1}{\sqrt{n}} \leq \|x\|_2$$

for all $x \in E$.

(2) Similar result holds for all $\ell_p^n, 1 \leq p \leq 2$.

(3) We only used the bounded volume ratio property of ℓ_1^n to get Kashin's result (geometry of ℓ_1^n is never used), so the same result holds for any space with good volume ratio property.

Another important consequence of VRT is the following splitting theorem due to Kashin.

Corollary 7.10 (Kashin's Splitting). *There exists an orthogonal decomposition $\mathbb{R}^n = E_1 \oplus E_2$ into $n/2$-dimensional subspaces such that both E_1 and E_2, as subspaces of ℓ_1^n, are within constant distance from Euclidean.*

Sketch of proof. Note that if a subspace E of dimension $\frac{n}{2}$ is uniformly distributed over the Grassmannian $G_{n,n/2}$, then so is its orthogonal complement $E^{\perp}$. Hence, both E and $E^{\perp}$ satisfies VRT with high probability, meaning that they are simultaneously almost Euclidean with high probability. ∎

7.3. Volume ratio theorem: global version. There are two branches of study in Banach spaces: local theory and global theory. "Local" usually refers to sections (subspaces) or projections (quotients), while "global" usually refers to the study of the entire space or intersections of different spaces. We have just presented the VRT in its local version. In this section, we will establish the global version.

Theorem 7.11 (Global VRT). *Let K be a symmetric convex body in $\mathbb{R}^n$. Suppose that $B_2^n \subset K$. Let $v(K)$ be the volume ratio. Then there exists $U \in O(n)$ such that*

$$\operatorname{diam}(K \cap UK) \leq (c \cdot v(K))^{\frac{1}{2}}.$$

Equivalently,

$$\max\{\|x\|_K, \|x\|_{UK}\} \geq (c \cdot v(K))^{-\frac{1}{2}} \|x\|_2.$$

Proof. (See Chapter 6 in [**?**]) Consider a Kashin's splitting $\mathbb{R}^n = E_1 \oplus E_2$ with

$$\dim E_1 = \dim E_2 = \frac{n}{2}$$

such that VRT holds for both E_1 and E_2. Then,

$$\|x\|_K \geq (c \cdot v(K))^{-\frac{1}{2}} \text{ for all } x \in E_1 \cup E_2.$$

Let P_1 and P_2 be orthogonal projections onto E_1 and E_2, respectively. Let $U = P_1 - P_2$. It is easy to see that $U \in O(n)$. Then,

$$\begin{aligned}
\max\{\|x\|_K, \|Ux\|_K\} &\geq \frac{1}{2}(\|x\|_K + \|x\|_{UK}) \\
&= \frac{1}{2}(\|P_1x + P_2x\|_K + \|P_1x - P_2x\|_K) \\
&\geq \max\{\|P_1x\|_K, \|P_2x\|_K\} \text{ (by triangle inequality)} \\
&\geq (cv(K))^{-\frac{1}{2}} \max\{\|P_1x\|_2, \|P_2x\|_2\} \\
&\geq cv(K)^{-\frac{1}{2}} \sqrt{\|P_1x\|_2^2 + \|P_2x\|_2^2} = cv(K)^{-\frac{1}{2}} \|x\|_2.
\end{aligned}$$

■

REMARKS. Note that $\max\{\|x\|_K, \|x\|_{UK}\} = \|x\|_{K \cap UK}$.

Part 2

Metric Entropy and Applications

8. Diameters of Projections of Convex Sets

Let K be a convex body in $\mathbb{R}^n$, and let P be a random projection onto a subspace of dimension k, that is, a projection onto a random subspace of dimension k. How does this projection change the shape of K? It is easy to see that a projection only catches the "extremal points" of K. Hence, a quantity affected by projections is the diameter of the body. In this section, we will examine how $\mathrm{diam}(PK)$ looks.

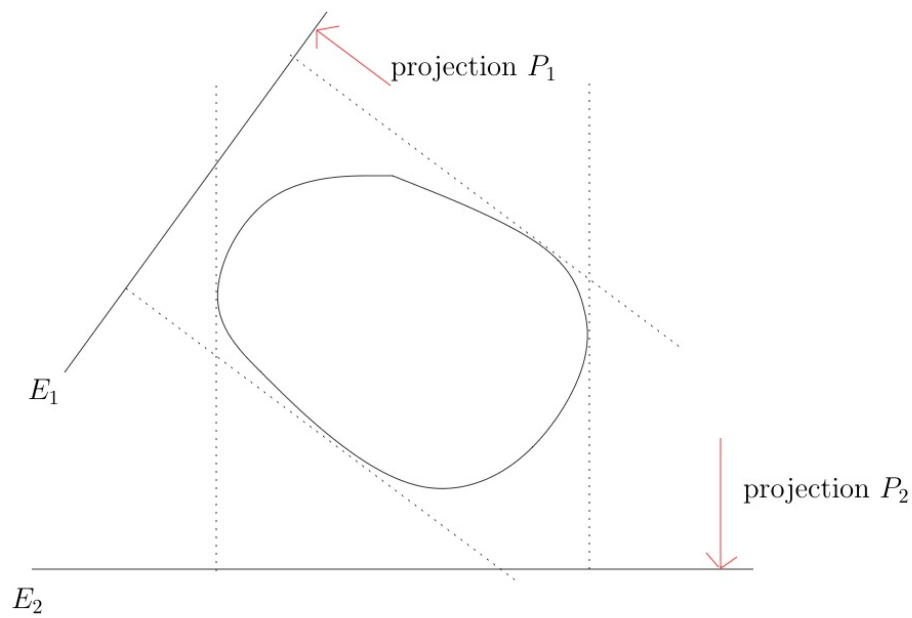

FIGURE 11. Diameters under Projections

EXAMPLE 8.1. (1) $K = [-1/2, 1/2]$. Then, by the estimate on the distance from a random vector to a random subspace,

$$\mathrm{diam}(PK) \sim \sqrt{\frac{k}{n}} \sim \sqrt{\frac{k}{n}}\,\mathrm{diam}(K).$$

(2) $K = B_2^n$. As any projection of a Euclidean ball is still Euclidean and does not change the diameter,

$$\mathrm{diam}(PK) = \mathrm{diam}(K).$$

¿From these two examples, we see two kinds of behavior of the diameter under projections: it "shrinks" by a factor of $\sqrt{n/k}$ in the first example, while it does not "shrink" at all in the second example. Surprisingly, those are essentially the only two possible types of projection effect on the diameter of a high-dimensional convex body. In order to explain this phenomenon, we first introduce the following notion called the "*mean width*" of a convex body.

DEFINITION 8.1. The *mean width* $M^*(K)$ of a symmetric convex body K in $\mathbb{R}^n$ is the expected value of the dual norm over the unit sphere (with respect to the normalized Haar measure). That is,

$$M^*(K) = \int_{S^{n-1}} \|x\|_{K^*} d\sigma(x) = \int_{s^{n-1}} \max_{y \in K} \langle x, y \rangle d\sigma(x).$$

REMARKS. (1) We can think of he mean width as the average "thickness" measured from the center of the convex body.

(2) The mean width of a symmetric convex body is the mean norm of its polar:

$$M^*(K) = M(K^\circ).$$

(3) Recall that

$$M(K) \sim \frac{\ell(K)}{\sqrt{n}},$$

where $\ell(K) = \mathbb{E}\|g\|_K$, and g is a standard Gaussian vector in $\mathbb{R}^n$. Hence,

$$M^*(K) = M(K^\circ) = \frac{\ell(K^\circ)}{\sqrt{n}} = \frac{\ell^*(K)}{\sqrt{n}},$$

where $\ell^*(K) = \mathbb{E}\|g\|_{K^\circ} = \mathbb{E}\max_{y \in K}\langle g, y \rangle$. The duality often makes the computation easier. We will see this in the following examples.

EXAMPLE 8.2 (Mean width of B_p^n). (1) Mean width of the cube.

$$M^*(B_\infty^n) = M(B_1^n) \sim \frac{\ell(B_1^n)}{\sqrt{n}} \sim \frac{n}{\sqrt{n}} = \sqrt{n} \sim \operatorname{diam}(B_\infty^n).$$

(2) Mean width of the crosspolytope.

$$M^*(B_1^n) = M(B_\infty^n) \sim \frac{\ell(B_\infty^n)}{\sqrt{n}} \sim \sqrt{\frac{\log n}{n}}.$$

Hence, the mean width of B_1^n is only logarithmically larger than the diameter of the inscribed ball, whereas the mean width of the cube is polynomially larger ($\sqrt{n}$) than its inscribed ball (as shown in the first example) . This is due to the contribution of very few vertices ($2n$ vertices instead of 2^n vertices as in the cube).

(3) For any $1 < p \le \infty$, let q be such that $1/p + 1/q = 1$. Then,

$$\begin{aligned} M^*(B_p^n) = M(B_q^n) \sim \frac{B_q^n}{\sqrt{n}} \sim n^{\frac{1}{q}-\frac{1}{2}} = n^{\frac{1}{2}-\frac{1}{p}} \\ \sim \begin{cases} \operatorname{diam}(\text{inscribed ball}), & 1 < p \le 2; \\ \operatorname{diam}(\text{circumscribed ball}), & p \ge 2. \end{cases} \end{aligned}$$

Now, if we apply the General Dvoretzky Theorem in the dual space, we get the following:

Theorem 8.3 (Dual Version of General Dvoretzky Theorem). *Let K be a (symmetric) convex body in $\mathbb{R}^n$ with* $\operatorname{diam}(K) \leq 1$. *Let*

$$k^* = c(\epsilon) M^*(K)^2 n.$$

Then a random projection P in $\mathbb{R}^n$ onto a k^*-dimensional subspace satisfies the following with exponentially high probability:

$$(1-\epsilon) M^*(K) \cdot P(B_2^n) \subseteq P(K) \subseteq (1+\epsilon) M^*(K) \cdot P(B_2^n).$$

Hence, the projection of K onto a k^*-dimensional subspace looks like a Euclidean ball of radius $M^*(K)$. Since the Euclidean ball is preserved by projections, any further projection will not change the diameter: PK will remain Euclidean for projections P onto subspaces of dimension $k \leq k^*$. But what happens for projections onto subspaces of dimension $k > k^*$? It turns out that $\operatorname{diam}(PK)$ will shrink by $\sqrt{n/k}$, just as in the case $K = [-1/2, 1/2]$.

Theorem 8.4 (Diameters under Random Projections, [**?**]). *Let K be a symmetric convex body in $\mathbb{R}^n$, and let P be a projection in $\mathbb{R}^n$ onto a random subspace of dimension k. Then, with probability at least $1 - e^{-k}$,*

$$\operatorname{diam}(PK) \leq C\Big(M^*(K) + \sqrt{\frac{k}{n}} \operatorname{diam}(K)\Big).$$

Proof. Without loss of generality, we may assume that $\operatorname{diam}(K) = 2$, i.e. $K \subseteq B_2^n$. Let $E \in G_{n,k}$ be a random subspace. Then, E can be viewed as the image of $\mathbb{R}^k$ under a random orthogonal transformation $U \in O(n)$. Therefore, the orthogonal projection P onto E can be viewed as the first k rows of U, and let us call it Q. Then $\operatorname{diam}(PK)$ is distributed identically with $\operatorname{diam}(QK)$. Note that

$$\operatorname{diam}(QK) = 2 \max_{x \in K} \|Qx\|_2,$$

and that for any $u \in \mathbb{R}^k$,

$$\|u\|_2 = \max_{z \in S^{k-1}} \langle u, z\rangle \leq 2 \max_{z \in \mathcal{N}} \langle u, z\rangle,$$

where $\mathcal{N}$ is a $\frac{1}{2}$-net of S^{k-1}. Choose $\mathcal{N}$ such that $|\mathcal{N}| \leq e^{ck}$.Then we have

$$\operatorname{diam}(QK) \leq 4 \max_{z \in \mathcal{N}} \max_{x \in K} \langle Qx, z\rangle.$$

Note that for fixed z and x,

$$\langle Qx, z\rangle = \langle x, Q^* z\rangle = \langle x, u\rangle,$$

where u is uniform random on S^{n-1}. This is because the adjoint Q^* is a random embedding of $\mathbb{R}^k$ into $\mathbb{R}^n$, which puts any fixed $z \in S^{k-1}$ uniformly distributed on S^{n-1}. Thus,

$$\mathbb{E} \max_{x \in K} \langle x, u\rangle = M^*(K).$$

It is easy to see the function $u \mapsto \max_{x\in K}\langle x, u\rangle$ is 1-Lipschitz on S^{n-1}. By concentration of measure on the sphere,

$$\Pr\left(\max_{x\in K}\langle x, u\rangle \geq M^*(K) + t\right) \leq e^{-t^2n/2} \; \forall t > 0.$$

Take union bound over $\mathcal{N}$, and we get

$$\Pr\left(\max_{z\in\mathcal{N}}\max_{x\in K}\langle x, u\rangle \geq M^*(K) + t\right) \leq |\mathcal{N}|e^{-t^2n/2} \leq e^{ck-t^2n/2} \leq e^{-k},$$

provided that $t = c\sqrt{k/n}$. Hence,

$$\Pr\left(\operatorname{diam}(QK) \geq C\Big(M^*(K) + \sqrt{\frac{k}{n}}\Big)\right) \leq e^{-k}.$$

■

Remarks. (1) If we let $\operatorname{diam}(K) = 1$ and $k^* := M^*(K)^2n$, then, with probability at least $1 - e^{-k}$,

$$\operatorname{diam}(PK) \leq \begin{cases} C\sqrt{\frac{k}{n}}, & k \geq k^*; \\ M^*(K), & k \leq k^*. \end{cases}$$

We see a "phase transition" at k^*: when $k > k^*$, the diameter of K shrinks by $\sqrt{n/k}$ under projection (as if K were an interval); when $k \leq k^*$, the diameter stabilizes since PK is already Euclidean.

(2) The result is sharp. See [**?**].

Corollary 8.5. *Let T be a linear operator on ℓ_2^n, and let P be the orthogonal projection onto a random subspace in the Grassmanian $G_{n,k}$. Then, with probability at least $1 - e^{-k}$,*

$$\|PT\| \leq C\Big(\frac{1}{\sqrt{n}}\|T\|_{HS} + \sqrt{\frac{k}{n}}\|T\|\Big).$$

Proof. We apply Theorem 8.4 for the ellipsoid $K := TB_2^n$. It is an exercise to check that $M^*(K) \approx \frac{1}{\sqrt{n}}\|T\|_{HS}$ and $\operatorname{diam}(K) = \|T\|$. ■

Remarks. There is a version of this corollary for projections onto random *coordinate* subspaces, see [M. Rudelson, R. Vershynin, Sampling from large matrices: an approach through geometric functional analysis, Journal of the ACM (2007), Art. 21, 19 pp].

Exercise 11. Get an estimate on the diameter under a random projection for

(1) $K = B_p^n, 1 < p \leq 2$.
(2) $K = B_p^n, 2 \leq p < \infty$.

9. Metric Entropy. Sudakov and Dual Sudakov Inequalities

One of the big open problems in geometric functional analysis is the duality conjecture. Recall that, given two convex bodies K and L, the *covering number* of K by L, denoted $N(K, L)$, is the minimum number of copies of L needed to cover K. The duality problem asks whether the covering number of K by L is equivalent to the covering number of L° by K°. Note that just for the dimension effect on the volume, we should expect a generic covering number exponential in the dimension. Hence, sometimes it is easier to work with $\log N(K, L)$, which we refer to as the *metric entropy* of K with respect to L. Now, we may phrase the duality problem as follows.

Conjecture 9.1 (Entropy duality). *There exist absolute constants $C, c > 0$ such that for every symmetric convex sets $K, L \subset \mathbb{R}^n$,*

$$c \log N(L^\circ, C \cdot K^\circ) \le \log N(K, L) \le C \log N(L^\circ, c \cdot K^\circ).$$

Remarks. (1) Clearly, for all K, L, the left hand side follows from the right hand side, and vice versa.
(2) It was recently solved for L being an ellipsoid. See [**?**].

There are two common ways to compute metric entropy:

(1) By volume ratio estimate

$$N(K, D) \le \frac{\operatorname{vol}\left(K + \frac{1}{2}D\right)}{\operatorname{vol}\left(\frac{1}{2}D\right)}.$$

However, this estimate can be very insensitive. One can see this through the example where K is a “lean sausage” and D a Euclidean ball.

(2) By the mean width. We will focus on this approach here.

Recall that $\ell^*(K) = \mathbb{E}\|g\|_{K^\circ} = \mathbb{E}\max_{x\in K}\langle g, x\rangle$. It turns out that the metric entropy and this Gaussian mean width are closely related.

Theorem 9.2 (Sudakov Inequalities). *For every symmetric convex set K in $\mathbb{R}^n$,*

(1) *Sudakov inequality:* $\sqrt{\log N(K, B_2^n)} \le C \cdot \ell^*(K)$;
(2) *Dual Sudakov inequality:* $\sqrt{\log N(B_2^n, K^\circ)} \le C \cdot \ell^*(K)$.

Remarks. (1) Note that the two inequalities are equivalent by the solved case of the duality conjecture. That is, when L is a Euclidean ball. See [**?**].
(2) More generally,

$$\sqrt{\log N(K, \epsilon B_2^n)} = \sqrt{\log N\left(\frac{1}{\epsilon}K, B_2^n\right)} \le C \cdot \frac{\ell^*(K)}{\epsilon},$$

and

$$\sqrt{\log N(B_2^n, \epsilon K^\circ)} \le C \cdot \frac{\ell^*(K)}{\epsilon}.$$

As the two inequalities implies each other, it suffices to prove only one of them. We will prove the dual Sudakov inequality. For reference, see Section 3.3 in [**?**]. We will imitate the proof of the volumetric bound for covering numbers. Recall that minimal covering is equivalent to maximal packing. Let $\{x_i, i = 1, \dots, N\}$ be a maximal 1-separated set in B_2^n in the norm induced by K°, then $\{x_i + K^\circ\}$ is a covering of B_2^n by maximality. Thus

$$N(B_2^n, K^\circ) \le N.$$

Moreover, the sets $x_i + K^\circ/2$ are disjoint (also from maximality), so they form a packing in the Gauss space $(\mathbb{R}^n, \gamma_n)$. Hence,

$$1 = \gamma_n(\mathbb{R}^n) \ge \gamma_n\left(B_2^n + \frac{1}{2}K^\circ\right) \ge \sum_{i=1}^{N} \gamma_n\left(x_i + \frac{1}{2}K^\circ\right).$$

The difficulty here is that the Gaussian measure is NOT translation invariant: the summand measures are not all the same. Fortunately, we have the following nice estimates for the Gaussian measure under translation and dilation.

Lemma 9.3 (Gaussian Measure of Translates). *For every symmetric set $T \subset \mathbb{R}^n$ and every $z \in \mathbb{R}^n$,*

$$\gamma_n(z + T) \ge e^{-\|z\|_2^2/2}\gamma_n(T).$$

Proof. First, note that $\gamma_n/\gamma_n(T)$ defines a probability measure on T. Hence,

$$\begin{aligned}
\gamma_n(z + T) &= (2\pi)^{-n/2} \int_{z+T} e^{-\|x\|_2^2/2} dx \\
&= (2\pi)^{-n/2} \int_{T} e^{-\|y+z\|_2^2/2} dy \\
&= e^{-\|z\|_2^2/2} \int_{T} e^{-\langle z,y\rangle} d\gamma_n(y) \\
&= e^{-\|z\|_2^2/2}\gamma_n(T) \int_{T} e^{-\langle z,y\rangle} d\gamma_n(y)/\gamma_n(T) \\
&\ge e^{-\|z\|_2^2/2}\gamma_n(T) e^{-\mathbb{E}\langle z,y\rangle} d\gamma_n(y) \quad \text{by Jensen's inequality} \\
&= e^{-\|z\|_2^2/2}\gamma_n(T) e^{-\langle z,\mathbb{E}y\rangle} d\gamma_n(y) \\
&= e^{-\|z\|_2^2/2}\gamma_n(T)
\end{aligned}$$

∎

Lemma 9.4 (Gaussian Measure of Dilations). *For any $t > 1$,*

$$\gamma_n\left(t\ell^*(K) \cdot K^\circ\right) \ge 1 - \frac{1}{t}.$$

PROOF. Note that $\ell^*(K) = \mathbb{E}\|g\|_{K^\circ}$. By Markov inequality,

$$\Pr\left(\|g\|_{K^\circ} > t\ell^*(K)\right) \le \frac{\mathbb{E}\|g\|_{K^\circ}}{t\ell^*(K)} = \frac{1}{t}.$$

Then,

$$\gamma_n\left(t\ell^*(K)\cdot K^\circ\right) = \Pr\left(g \in t\ell^*(K)\cdot K^\circ\right) = \Pr\left(\|g\|_{K^\circ} \le t\ell^*(K)\right) \le 1 - \frac{1}{t}.$$

Let $t = 1/2$. Then we have

$$\gamma_n(2\ell^*(K)\cdot K^\circ) \ge \frac{1}{2}. \tag{11}$$

■

Now we are in a position to prove dual Sudakov inequality.

PROOF OF DUAL SUDAKOV. Let $a = \ell^*(K)/4$. Note that

$$N := N(B_2^n, K^\circ) = N(aB_2^n, aK^\circ).$$

Then there exists $\{x_i, i = 1, \dots, N\}$ such that $x_i + (a/2)K^\circ$ are pairwise disjoint. Then, by Lemma 9.3 and inequality (11),

$$\begin{aligned}1 &\ge \sum_{i=1}^N \gamma_n(x_i + \frac{a}{2}K^\circ) \ge \sum_{i=1}^N e^{-\|x_i\|_2^2/2}\gamma_n(\frac{a}{2}K^\circ)\\ &\ge \sum_{i=1}^N e^{-a^2/2}\cdot\frac{1}{2} \ge \frac{1}{2}e^{-a^2/2}N.\end{aligned}$$

Hence, $N \le 2e^{a^2/2}$, and thus $\sqrt{\log N} \le C\ell^*(K)$. ■

Next, we will prove a weak form of the duality of metric entropy, which is sufficient for the equivalence of Sudakov and dual Sudakov inequalities.

LEMMA 9.5 (Weak Entropy Duality). *For any symmetric convex set K in $\mathbb{R}^n$,*

$$\sup_{\epsilon>0} \epsilon\sqrt{\log N(K, \epsilon B_2^n)} \le 8 \sup_{\epsilon>0} \epsilon\sqrt{\log N(B_2^n, \epsilon K^\circ)}.$$

We will need a couple of easy facts in the proof of this lemma.

PROPOSITION 9.6. *For any $t, s > 0$ and any symmetric convex body K in $\mathbb{R}^n$,*

$$tK \cap sK^\circ \subseteq \sqrt{ts}B_2^n.$$

PROOF. For any $x \in tK \cap sK^\circ$, $\|x\|_2^2 = \langle x, x\rangle \le \|x\|_K\|x\|_{K^\circ} \le ts$. ■

PROPOSITION 9.7. *Let $K, L \subset \mathbb{R}^n$. Suppose that K is symmetric (i.e.$K = -K$). Then, $N(K, L) \ge N(K, 2K \cap L)$.*

PROOF. Let $N = N(K, L)$. Then, there exist $\{x_i\}_{i=1}^N \subset K$ such that

$$K \subset \bigcup_{i=1}^N x_i + L.$$

Then, for any $y \in K$, there exists some $i \leq N$ such that $y - x_i \in L$. Note that by symmetry of K, $y - x_i \in K + K = 2K$. Hence,

$$K \subset \left(\bigcup_{i=1}^N x_i \right) + (2K \cap L),$$

and thus $N(K, 2K \cap L) \leq N$. ∎

PROOF OF DUAL SUDAKOV INEQUALITY. By Proposition 9.6 and Proposition 9.7,

$$\begin{aligned}
N(K, \epsilon B_2^n) &\leq N\left(K, 2K \cap \frac{\epsilon^2}{2} K^\circ\right) \\
&\leq N\left(K, \frac{\epsilon^2}{2}\right) \\
&\leq N(K, 2\epsilon B_2^n) \cdot N\left(2\epsilon B_2^n, \frac{\epsilon^2}{2} K^\circ\right) \\
&= N(K, 2\epsilon B_2^n) \cdot N\left(B_2^n, \frac{\epsilon}{4} K^\circ\right).
\end{aligned}$$

Taking log, square root, and multiplying both sides by ϵ, with the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we get

$$\epsilon \sqrt{\log N(K, \epsilon B_2^n)} \leq \epsilon \sqrt{\log N(K, \frac{\epsilon}{2} B_2^n)} + \sup_{\epsilon > 0} \epsilon \sqrt{\log N(B_2^n, \frac{\epsilon}{4})}.$$

Let $f(\epsilon) = \epsilon \sqrt{\log N(K, \epsilon B_2^n)}$ and $M = \sup_{\epsilon > 0} \epsilon \sqrt{\log N(B_2^n, \epsilon K^\circ)}$. Then,

$$f(\epsilon) \leq \frac{1}{2} f(2\epsilon) + 4M.$$

We leave it as an exercise to check that $\sup_{\epsilon > 0} f(\epsilon) \leq 8M$. ∎

EXERCISE 12. If a function satisfies

(1) $f(\epsilon) \leq \frac{1}{2} f(2\epsilon) + 4M$;
(2) $\lim_{\epsilon \to \infty} f(\epsilon) = 0$,

then $\sup_{\epsilon > 0} f(\epsilon) \leq 8M$.

EXAMPLE 9.8. $K = B_1^n$. Clearly, $N(K, \epsilon B_2^n) \geq 2n$ for all $\epsilon \leq 1$ just to cover the vertices. By Sudakov inequalities,

$$N(B_1^n, \epsilon B_2^n) \leq e^{c\ell^*(K)^2/\epsilon^2} \leq n^{c/\epsilon^2},$$

as $\ell^*(K) \sim \sqrt{\log n}$. The significance is this upper bound is polynomial in n, which matches the lower bound.

Exercise 13. Prove that for any polytope K in $\mathbb{R}^d$ with n vertices can be covered by $poly_\epsilon(n)$ Euclidean balls of diameter $\leq \epsilon \cdot \mathrm{diam}(K)$. (Hint: Realize K as some projection of B_1^n from $\mathbb{R}^n$ to $\mathbb{R}^d$.)

Remarks. See Section 3.3 in [**?**] for a different proof under a reformulation for Gaussian processes.

10. Low M^*-estimate. ℓ-position

In the previous section, we saw how diameter behaves under a random orthogonal projection. In this section, we will look at the dual case. We will see how diameter behaves in a random section.

Theorem 10.1 (Low M^*-estimate; Theorem 18.2 in [**?**]; [**?**]; [**?**]). *Let K be a symmetric convex body in $\mathbb{R}^n$, and let $E \in G_{n,n-k}$ be a random subspace of codimension k. Then, with probability at least $1 - e^{-k}$,*

$$\mathrm{diam}(K \cap E) \leq C\sqrt{\frac{n}{k}} M^*(K).$$

Remarks. (1) The diameter estimate was first given by Milman with linear dependence on n, and later improved by Pajor and Tomczak ([**?**]).
(2) As $\ell^*(K) \sim \sqrt{n} M^*(K)$, we have a dimension-free version

$$\mathrm{diam}(K \cap E) \leq \frac{C}{\sqrt{k}} \ell^*(K).$$

Proof. We will follow a standard discretization argument.
Step 1:Discretizing K using Sudakov inequality.
Let

$$t = c\sqrt{\frac{n}{k}} M^*(K).$$

For computational purposes, we will take $c = \sqrt{10}$. By Sudakov inequality, there exists a t-net $\mathcal{N}$ of K in Euclidean metric of cardinality

$$|\mathcal{N}| \leq \exp\left(C\frac{\ell^*(K)^2}{t^2}\right) \sim \exp\left(\frac{nM^*(K)^2}{t^2}\right) = e^{k/10}.$$

Step 2: Fixed vector: diameter of random projection.
Let $x \in K \cap E$. Then there exists $y \in \mathcal{N}$ with $\|x - y\|_2 \leq t$. Consider the orthogonal projection P onto $E^\perp \in G_{n,k}$. Then $x \in \ker P$, and thus

$$\|Py\|_2 = \|P(x - y)\|_2.$$

Moreover, since K is convex and symmetric,

$$x - y \in 2K.$$

Hence,

$$x - y \in tB_2^n \cap 2K.$$

By Theorem 8.4, we have that, with probability at least $1 - e^{-k}$,

$$\operatorname{diam}(P(tB_2^n \cap 2K)) \le C\left(M^*(K) + t\sqrt{\frac{k}{n}}\right) \le CM^*(K),$$

using our prescribed value for t in the beginning of Step 1. Choose a realization of P that satisfies this. Then, for all $x \in K \cap E$, there exists $y \in \mathcal{N}$ such that

$$\|Py\|_2 \le \|P(x-y)\|_2 \le \operatorname{diam}(P(tB_2^n \cap K)) \le CM^*(K). \tag{12}$$

Step 3: Norm under random projection and union bound.
Recall that for any fixed $y \in \mathbb{R}^n$ and for a random projection onto a k-dimensional subspace,

$$\|Py\|_2 \ge c\sqrt{\frac{k}{n}}\|y\|_2$$

with probability at least $1 - e^{-k}$. Hence, by taking union bound over $\mathcal{N}$, we get

$$\|Py\|_2 \ge c\sqrt{\frac{k}{n}}\|y\|_2 \tag{13}$$

for all $y \in \mathcal{N}$ with probability at least $1 - |\mathcal{N}| \cdot 2e^{-k} > 1 - e^{-ck}$, using the bound on $|\mathcal{N}|$ in the first step.
Step 4: Approximation.
Combine the upper and lower bounds from (12) and (13), and we get that for all $y \in \mathcal{N}$, $\|y\|_2 \le C\sqrt{n}kM^*(K)$. Hence, for all $x \in K$,

$$\|x\|_2 \le C\sqrt{\frac{n}{k}}M^*(K) + t \le C\sqrt{\frac{n}{k}}M^*(K)$$

by our choice of t. ■

As an example, we now apply this low M^*-estimate to examine Euclidean subspaces of ℓ_p^n of arbitrary dimensions.

EXAMPLE 10.2. $B_p^n, 1 < p \le 2$.
Note that

$$M^*(B_p^n) \approx \text{diameter of its inscribed ball} \sim n^{\frac{1}{2}-\frac{1}{p}}.$$

Consider $K = n^{\frac{1}{2}-\frac{1}{p}}B_p^n$ so that B_2^n is the inscribed ball. Then

$$M^*(K) \sim \text{ constant}.$$

By low M^*-estimate, for a random $E \in G_{n,n/2}$,

$$\operatorname{diam}(K \cap E) \sim \text{ constant}.$$

So the random section $K \cap E$ is Euclidean.

COROLLARY 10.3. *(Euclidean subspaces of ℓ_p^n, $1 < p \le 2$) A random subspace $E \in G_{n,n-k}$ of ℓ_p^n, $1 < p \le 2$ satisfies*

$$d(E, \ell_p^n) \le C_p \sqrt{\frac{n}{k}}$$

with probability at least $1 - e^{-k}$.

REMARKS. (1) This is not sharp for $k = 1$ (small codimensional subspaces):

$$d(E, \ell_2^{n-1}) \approx d(\ell_p^n, \ell^n) \sim n^{\frac{1}{p} - \frac{1}{2}} \ll C_p \sqrt{n}.$$

(2) A similar result follows from the volume ratio theorem since

$$v(\ell_p^n) \sim \text{ constant}$$

for $1 \le p \le 2$. However, the volume ratio theorem gives an exponential dependence of $d(E, \ell_2^{n-k})$ on the aspect ratio:

$$d(E, \ell_2^{n-k}) \le C^{\frac{n}{k}},$$

whereas the low M^*-estimate gives a polynomial dependence.

(3) For new developments on this, see [**?**] and [**?**], where asymptotic formula for $\mathbb{E}\,\mathrm{diam}(K \cap E)$ are given in terms of $M^*(K \cap rB_2^n)$, where $r = r(k/n)$.

EXAMPLE 10.4. B_1^n.
Consider $K = \sqrt{n} B_1^n$. We have seen before that

$$M^*(K) \sim \sqrt{\log n}.$$

Hence, the low M^*-estimate implies

$$d(E, \ell_2^{n-k}) \le C \sqrt{\frac{k}{n} \log n}.$$

REMARKS. A more accurate estimate, with $\log \frac{n}{k}$ factor instead of $\log n$, was obtained in [**?**]. We will quote it here without proof.

THEOREM 10.5 (Diameter of Euclidean sections of B_1^n; [**?**]). *A random subspace $E \in G_{n,n-k}$ of ℓ_1^n satisfies the following with probability at least $1 - e^{-ck}$:*

$$\mathrm{diam}(E \cap \sqrt{n} B_1^n) \le C \sqrt{\frac{n}{k} \log \frac{n}{k}}.$$

A *position* of a convex body is a linear transformation of the body. For example, *John's position* of a convex body K is the linear image TK such that the ellipsoid of maximal volume in K is B_2^n. It is the position that minimizes the volume ratio of K. In general, choice of a position of a convex body corresponds to choice of a coordinate system to work with. That is, we are choosing a Euclidean structure on the linear space that contains K. This is extremely convenient especially when the initial setup is in some abstract linear space without a coordinate system. In this section, we will discuss an important position of convex bodies called the *ℓ-position.*

DEFINITION 10.1 (ℓ-position; [**?**]). An *ℓ-position* of a convex body K minimizes the product $M(TK)M^*(TK)$ over all positions TK of K.

Recall that $M(K)M^*(K) \geq 1$ for any convex body K. However, the converse does not hold: $M(K)M^*(K)$ can be arbitrarily large.

EXERCISE 14. Show that $M(K)M^*(K)$ can be arbitrarily large in $\mathbb{R}^n$.

The good news is that if we are allowed position K, then we have an upper bound for MM^*. The following result is given without proof. See [**?**] and Chapter 15 in [**?**].

THEOREM 10.6 (Upper Bound for MM^*). *For every symmetric convex body K in $\mathbb{R}^n$, there exists a position TK such that*

$$M(TK)M^*(TK) \leq C \log 2d(K, B_2^n) \leq C \log n,$$

where the last inequality is immediate from John's theorem.

REMARKS. (1) The result would fail if the convex body is non-symmetric. It is still an open problem when K is non-symmetric. The best known bound is $O^*(n^{4/3})$ by Rudelson [**?**].
(2) The result of the theorem is sharp. We leave this as an exercise. Hence, we also call the position in the theorem an *ℓ-position of* K.

EXERCISE 15. Prove sharpness of this theorem. That is, find some K such that $M(K)M^*(K) \sim \log n$.
Hint: For $K = B_1^n$, we have

$$M(K) \sim \sqrt{n} \text{ and } M^*(K) \sim \sqrt{\frac{\log n}{n}}$$

so that $MM^* \sim \sqrt{\log n}$. If there could be a $\sqrt{\log n}$ factor in both $M(K)$ and $M^*(K)$, then we would be done. Is there a way to modify B_1^n so this can happen?

11. Quotient of Subspace Theorem

In this section, we will see that every symmetric convex body in a finite dimensional normed space contains a Euclidean structure. We have already seen that convex bodies such as B_1^n has large Euclidean sections, while convex bodies such as B_∞^n has large Euclidean projections. It is then natural to think that, given any symmetric convex body, we could end up with some large Euclidean structure if we compose these two operations. This idea is confirmed by the *Quotient of Subspace Theorem* (QS Theorem for short). We will follow [**?**].

Theorem 11.1 (Quotient of Subspace theorem). *Let K be a symmetric convex body in $\mathbb{R}^n$, and let $\delta \in (0,1)$. Then there exist subspaces $F \subset E \subset \mathbb{R}^n$ with $k = \dim F \geq (1-\delta)n$ such that*

$$d\left(P_F(K \cap E), \ell_2^k\right) \leq \frac{C}{\delta} \log \frac{1}{\delta}.$$

Equivalently, let X be an n-dimensional normed space. Let $\delta \in (0,1)$. Then there exist subspaces $H \subset E \subset X$ with $k = \dim(E/H) \geq (1-\delta)n$ such that

$$d\left(E/H, \ell_2^k\right) \leq \frac{C}{\delta} \log \frac{1}{\delta}.$$

Let us sketch the idea of the proof first. Recall that a quotient of a subspace and a subspace of a quotient are related by duality, and that $M^*(K^\circ) = M(K)$. Then,

$$K \cap E \subseteq M^*(K) \cdot B_2^n \text{ iff } P_E(K) \supseteq \frac{1}{M^*(K)} \cdot B_2^n.$$

Similarly, if we take a further section of the projection,

$$P_E(K) \cap F \subseteq M(K) \cdot B_2^n.$$

Hence,

$$d(P_E(K) \cap F, \ell_2^k) \leq M(K)M^*(K) \leq \log d(K, B_2^n).$$

This means that the distance from Euclidean drops logarithmetically after we take a composition of quotient and subspace operation. We then may hope that the distance can drop to constant order after a few iterations.

Proof. We will proceed by steps.
Step 1: Double application of low M^*-estimate.
By low M^*-estimate, there exists a subspace E with $\dim E = (1-\delta)n$ such that

$$K \cap E \subseteq \frac{C}{\sqrt{\delta}} M^*(K) \cdot B_2^n \cap E.$$

By duality,

$$P_E(K^\circ) \supseteq c\sqrt{\delta} \frac{1}{M^*(K)} \cdot B_2^n \cap E.$$

Now, apply low M^*-estimate again, and we get a further subspace $F \subset E$ with $\dim F = (1-\delta)\dim E = (1-\delta)^2 n$ such that

$$P_E(K^\circ) \cap F \subseteq \frac{C}{\sqrt{\delta}} M^*(P_E(K^\circ)) \cdot B_2^n \cap F.$$

Note that by duality, $M^*(P_E(K^\circ)) = M(K \cap E)$, and we leave it an exercise to show that $M(K \cap E) \leq M(K)$. Thus,

$$c\sqrt{\delta} \frac{1}{M^*(K)} \cdot B_2^n \cap F \subseteq P_E(K^\circ) \cap F \subseteq \frac{C}{\sqrt{\delta}} M(K) \cdot B_2^n \cap F.$$

Hence,

$$d(P_E(K^\circ) \cap F, \ell_2^k) \leq \frac{C}{\delta} M(K)M^*(K) \leq \frac{C}{\delta} \log 2d(K, \ell_2^n)$$

by MM^*-estimate. Recall that $d(X,Y) = d(X^*, Y^*)$ for any Banach spaces X and Y. Hence, we have shown that

$$d(P_F(K \cap E), \ell_2^k) \leq \frac{C}{\delta} \log 2d(K, \ell_2^n).$$

Step 2: Iteration.
Consider

$$QS(K) := \{ \text{ projections of sections of } K\}.$$

Define $f : (0,1) \to \mathbb{R}^+$ by

$$f(\alpha) = \inf\{d(L, \ell_2^k : L \in QS(K), k = \dim L \geq \alpha n\}.$$

We have shown in the first step that f satisfies the following recursive relation:

$$f((1-\delta)^2\alpha) \leq \frac{C}{\delta} \log(2f(\alpha)), \;\; \forall \alpha \in (0,1) \; \forall \delta \in (0,1).$$

It can then be shown that functions satisfying this must satisfy

$$f(\alpha) \leq \frac{C}{1-\alpha} \log \frac{1}{1-\alpha}.$$

See Section 5.9 in [**?**] for details. ∎

Exercise 16. Prove that $M(K \cap E) \leq M(K)$ for any symmetric convex body K in $\mathbb{R}^n$ and any subspace E.

Part 3

Geometric Inequalities

In the last part of this course, we will see a series of geometric inequalities involving volumes and their applications in geometric functional analysis. For example, isoperimetric inequalities and concentration of measure inequalities can be recovered from Brunn-Minkowski inequality. Moreover, we will learn a new position of convex bodies called the notion in establishing the inverse inequalities to the classic ones such as Brunn-Minkowski and Santalo inequalities. We will also revisit the entropy duality problem and the quotient of subspace theorem, as a result of these geometric inequalities.

12. Brunn-Minkowski Inequality

12.1. Brunn-Minkowski inequality. Brunn-Minkowski inequality is one of the most fundamental results in convex geometry. As we have seen all along this course, the correspondence between convex geometry and functional analysis produces better understanding of the whole picture. Hence, we should expect that Brunn-Minkowski inequality has influential consequences in geometric functional analysis. It turns out that the results on concentration of measure in various spaces we saw in the beginning of the course can actually be derived from Brunn-Minkowski inequality. Although it has profound consequences, the form of the inequality is extremely simple.

Theorem 12.1 (Brunn-Minkowski inequality). *For any measurable sets $A, B \subset \mathbb{R}^n$, we have:*

(1) *(additive form)* $\operatorname{vol}(A+B)^{1/n} \geq \operatorname{vol}(A)^{1/n} + \operatorname{vol}(B)^{1/n}$;
(2) *(multiplicative form)* $\operatorname{vol}(\lambda A + (1-\lambda)B) \geq \operatorname{vol}(A)^{\lambda} \operatorname{vol}(B)^{1-\lambda}$.

Moreover, equality holds iff A and B are homothetic.

Remarks. (1) The two versions are equivalent. (1) implies (2) by replacing A by λA and B by $(1-\lambda B)$ and then applying the geometric-arithmetic inequality. For the other direction, we may apply (2) with

$$A' = \frac{A}{\operatorname{vol}(A)^{\frac{1}{n}}}, \quad B' = \frac{B}{\operatorname{vol}(B)^{\frac{1}{n}}}, \quad \lambda = \frac{\operatorname{vol}(A)^{\frac{1}{n}}}{\operatorname{vol}(A)^{\frac{1}{n}} + \operatorname{vol}(B)^{\frac{1}{n}}}.$$

(2) Recall that a *log-concave measure* μ is a measure that satisfies

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^{\lambda} \mu(B)^{1-\lambda}.$$

By the equivalence of these the versions of Brunn-Minkowski inequality, we see that any log-concave measure, not just the Lebesgue measure on $\mathbb{R}^n$, satisfies Brunn-Minkowski inequality.

12.2. Prekopa-Leindler inequality. Brunn-Minkowski inequality is a special case of Prekopa-Leindler inequality, which can be viewed as a reverse Hölder inequality.

LEMMA 12.2 (Prekopa-Leindler inequality). *Let $f, g, h : \mathbb{R}^n \to [0, \infty)$ be measurable functions, and let $\lambda \in (0, 1)$. Assume that*

$$h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda} \ \forall x, y \in \mathbb{R}^n \ . \tag{14}$$

Then,

$$\int_{\mathbb{R}^n} h \geq \left(\int_{\mathbb{R}^n} f \right)^\lambda \left(\int_{\mathbb{R}^n} g \right)^{1-\lambda} .$$

REMARKS. Prekopa-Leindler inequality immdiately implies Brunn-Minkowski inequality by taking indicator functions $f = 1_A$, $g = 1_B$, and $h = 1_{\lambda A + (1-\lambda)B}$. Hence, all we need to do is to prove this functional form of Brunn-Minkowski inequality.

PROOF OF PREKOPA-LEINDLER. We first use the one-dimensional Brunn-Minkowski inequality to prove the one-dimensional Prekopa-Leindler inequality. Then we will proceed by induction on the dimension.
Step 1: One-dimensional Brunn-Minkowski.
We will show that $|A+B| \geq |A| + |B|$, where $|\cdot|$ denotes the Lebesgue measure on the real line. Without loss of generality, assume A, B are compact. As the Lebesgue measure is translation invariant, we can also assume that

$$\sup A = 0 = \inf B.$$

Then,

$$A \subset A + B \text{ and } B \subset A + B.$$

Hence $|A+B| \geq |A| + |B|$.
Step 2:One-dimensional Prekopa-Leindler.
We will reduce the integral to level sets. Note that

$$\int f = \int_0^\infty |\{f \geq a\}| da. \tag{15}$$

It is easy to check that

$$\{h \geq a\} \supseteq \lambda\{f \geq a\} + (1-\lambda)\{g \geq a\}.$$

By one-dimensional Brunn-Minkowski, we have

$$|\{h \geq a\}| \geq \lambda |\{f \geq a\}| + (1-\lambda)|\{g \geq a\}|.$$

Integrate both sides over all $a > 0$. By (15), we have

$$\int h \geq \lambda \int f + (1-\lambda) \int g.$$

Then, apply arithmetic-geometric mean inequality, and we get

$$\int h \geq \left(\int f \right)^\lambda \left(\int g \right)^{1-\lambda} .$$

Step 3: Induction on dimension.
Assume that the lemma holds in $\mathbb{R}^{n-1}$. For $t \in \mathbb{R}$, consider the functions

$$f_t, g_t, h_t : \mathbb{R}^{n-1} \to \mathbb{R}$$

defined as

$$f_t(x) = f(t, x)$$

and similarly for g_t and h_t. Suppose that

$$t = \lambda t_1 + (1 - \lambda)t_2.$$

Then by (14),

$$h_t(\lambda x + (1 - \lambda)y) \geq f_{t_1}(x)^\lambda g_{t_2}(y)^{1-\lambda}.$$

for all $x, y \in \mathbb{R}^{n-1}$. By induction hypothesis,

$$\int_{\mathbb{R}^{n-1}} h_t \geq \left(\int_{\mathbb{R}^{n-1}} f_{t_1}\right)^\lambda \left(\int_{\mathbb{R}^{n-1}} g_{t_2}\right)^{1-\lambda}.$$

Now, apply the one-dimensional Prekopa-Leindler, and we get

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^{n-1}} h_t\right) \geq \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}^{n-1}} f_{t_1}\right)\right)^\lambda \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}^{n-1}} g_{t_2}\right)\right)^{1-\lambda}.$$

This completes the proof. ∎

As we said earlier, Brunn-Minkowski inequality can be obtained as soon as the Prekopa-Leindler inequality is proved. In the next section, we will see various consequences of Brunn-Minkowski inequality in not only convex geometry, but also Banach spaces.

13. Applications of Brunn-Minkowski Inequality

13.1. Brunn's Principle.

THEOREM 13.1. *Let K be a convex body in $\mathbb{R}^n$ and $u \in \mathbb{R}^n$. Consider the hyperplanes*

$$H_t = \{x \in \mathbb{R}^n : \langle x, u\rangle = t\}.$$

Then the function

$$t \mapsto \operatorname{vol}(K \cap H_t)^{\frac{1}{n-1}}$$

is concave on its support.

REMARKS. It is easy to check that if K is a cone, then this function will be linear.

PROOF. Consider the sections $K_t = K \cap H_t$. Convexity implies that for any $r, s \in \mathbb{R}$ and for any $\lambda \in (0, 1)$,

$$K_{\lambda r+(1-\lambda)s} \supseteq \lambda K_r + (1 - \lambda)K_s.$$

By Brunn-Minkowski inequality, we have

$$\operatorname{vol}(K_{\lambda r+(1-\lambda)s})^{\frac{1}{n-1}} \geq \lambda \operatorname{vol}(K_r)^{\frac{1}{n-1}} + (1 - \lambda) \operatorname{vol}(K_s)^{\frac{1}{n-1}},$$

which is exacly what we want. ∎

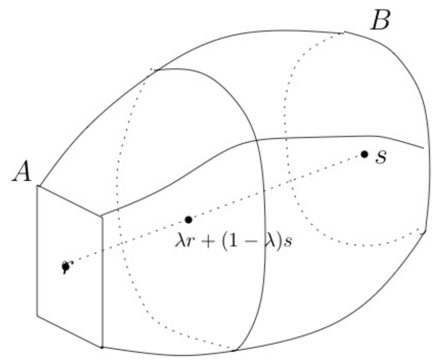

FIGURE 12. Brunn's Principle

REMARKS. (1) One can deduce Brunn-Minkowski inequality for *convex bodies* from Brunn's principle. Note that we state it for all measurable sets in Theorem 12.1.
We can embed A and B in $\mathbb{R}^{n+1}$ so that A is at position $t = 0$ and B at position $t = 1$ for the additional dimension t. See Figure 12. Then the section of the convex hull of $A \cup B$ at $t = 1-\lambda$ corresponds to the sets $\lambda A + (1-\lambda)B$. Then Brunn-Minkowski inequality is a direct consequence of Brunn's principle. We leave the details to the reader.
(2) Brunn's principle can also be derived from a direct geometric argument through Steiner's symmetrization. See Theorem 3.1 in [**?**].

13.2. Isoperimetric inequality in $\mathbb{R}^n$. Brunn-Minkowski inequality also implies the isoperimetric inequality in a very simple way.

Let A be a set in $\mathbb{R}^n$. Recall that the *surface area* $\mathrm{vol}(\partial A)$ of A is defined by

$$\mathrm{vol}(\partial A) = \lim_{\epsilon \to 0} \frac{\mathrm{vol}(A + \epsilon B_2^n) - \mathrm{vol}(A)}{\epsilon}. \tag{16}$$

THEOREM 13.2 (Isoperimetric inequality in $\mathbb{R}^n$). *Among all bodies of a given volume in $\mathbb{R}^n$, Euclidean balls have the least surface area.*

PROOF. Consider $A \subset \mathbb{R}^n$ and a Euclidean ball B with $\mathrm{vol}(A) = \mathrm{vol}(B)$. We want to show that

$$\mathrm{vol}(\partial A) \geq \mathrm{vol}(\partial B).$$

By Brunn-Minkowski inequality, for any $\epsilon > 0$,

$$\begin{aligned} \mathrm{vol}(A + \epsilon B_2^n)^{\frac{1}{n}} &\geq \mathrm{vol}(A)^{\frac{1}{n}} + \mathrm{vol}(\epsilon B_2^n)^{\frac{1}{n}} \\ &= \mathrm{vol}(B)^{\frac{1}{n}} + \mathrm{vol}(\epsilon B_2^n)^{\frac{1}{n}} \\ &= \mathrm{vol}(B + \epsilon B_2^n)^{\frac{1}{n}}, \end{aligned}$$

since B and ϵB_2^n are homothetic. Then use surface area formula (16) and take limit as $\epsilon \to 0$. ∎

Remarks. It is still not known how to apply Brunn-Minkowski to get isoperimetric inquality on the sphere, which is more general situation as spheres locally look like $\mathbb{R}^n$. The difficulty on the sphere is that subsets may "curve back" to wrap itself. Recall that in the proof of the one-dimensional Brunn-Minkowski, we first process the two sets A, B so that they are apart from each other. This causes a problem on the sphere: it is not always possible to separate two sets.

13.3. Concentration of measure in the ball, sphere and Gauss space. We know that isoperimetry on the sphere implies concentration of measure on the sphere early in this course, but we do not know how to get isoperimetry on the sphere from Brunn-Minkowski. Hence, in an effort to go from Brunn-Minkowski inequality to the concentration of measure on the sphere, we will bypass isoperimetry and take a direct path. Sometimes this is called "*approximate isoperimetry.*"

13.3.1. *Concentration of measure in the ball.* Consider B_2^n equipped with the normalized Lebesgue measure

$$\mu(A) = \frac{\text{vol}(A)}{\text{vol}(B_2^n)}.$$

Let

$$A_\epsilon := \{x \in B_2^n : d(x, A) \le \epsilon\}$$

denote the ϵ-extension of the set A.

Theorem 13.3 (Concentration in the Ball). *Let $A \subset B_2^n$ be a measurable set. Then*

$$\mu(A_\epsilon) \ge 1 - \frac{C}{\mu(A)} e^{-c\epsilon^2 n}$$

for any $\epsilon > 0$, where C, c are absolute constants. In particular, if $\mu(A) \ge 1/2$, then

$$\mu(A_\epsilon) \ge 1 - 2Ce^{-c\epsilon^2 n}.$$

Proof. Consider

$$B = (A_\epsilon)^c = \{x \in B_2^n : d(x, A) > \epsilon\}.$$

Our goal is to show that

$$\mu(A)\mu(B) \le Ce^{c\epsilon^2 n}.$$

One can check that μ is a log-concave measure. By Brunn-Minkowski inequality for log-concave measures, we have

$$\sqrt{\mu(A)\mu(B)} \le \mu\left(\frac{A+B}{2}\right).$$

Hence, it suffices to show that

$$\mu\left(\frac{A+B}{2}\right) \le Ce^{-c\epsilon^2 n}.$$

This will be immediate from the following claim.

Claim: $\frac{A+B}{2} \subset \sqrt{1-\frac{\epsilon^2}{4}}B_2^n$.

Let $x \in A$ and $y \in B$. Then $d(x,y) > \epsilon$. We need to estimate $\|\frac{x+y}{2}\|$. By the parallelogram law,

$$\begin{aligned}\|\frac{x+y}{2}\| &= \frac{1}{2}\sqrt{2(\|x\|^2+\|y\|^2)-\|x-y\|^2}\\ &\leq \sqrt{1-\frac{\epsilon^2}{4}},\end{aligned}$$

as $\|x\| \leq 1, \|y\| \leq 1$, and $\|x-y\| > \epsilon$. This finishes the proof of the claim. Hence,

$$\begin{aligned}\mu\left(\frac{A+B}{2}\right) &\leq \mu\left(\sqrt{1-\frac{\epsilon^2}{4}}B_2^n\right)\\ &= \left(1-\frac{\epsilon^2}{4}\right)^{n/2} \leq e^{-\epsilon^2 n/8}.\end{aligned}$$

This completes the proof. ∎

13.3.2. *Concentration of measure on the sphere.* Concentration on the sphere follows from concentration of measure in the ball. We will leave the proof to the reader.

Corollary 13.4 (Concentration on the Sphere). *Let $A \subset S^{n-1}$ be measurable, and let σ denote the canonical probability measure over S^{n-1}. Then*

$$\sigma(A_\epsilon) \geq 1 - \frac{C}{\sigma(A)}e^{-c\epsilon^2 n}$$

Exercise 17. Prove this corollary.

13.3.3. *Concentration of measure in Gauss space.* Recall that Gauss space is just $\mathbb{R}^n$ equipped with the standard n-dimensional Gaussian measure γ. We will deduce concentration of measure in Gauss space directly from Prekopa-Leindler inequality.

Theorem 13.5 (Concentration of Measure in Gauss Space; [?]). *Let $A \subset \mathbb{R}^n$ be a measurable set. Let $g \in \mathbb{R}^n$ be a standard Gaussian vector. Then*

$$\mathbb{E}\exp\left(\frac{d(g,A)^2}{4}\right) \leq \frac{1}{\gamma(A)}.$$

In particular,

$$\gamma(A_t) \geq 1 - \frac{e^{-t^2/4}}{\gamma(A)}.$$

Proof. The "in particular" part follows from Markov inequality:

$$\gamma(A_t) = \gamma\{x : d(x,A) > t\} \leq \frac{1}{\gamma(A)} \cdot e^{-t^2/4}.$$

For the main result, we apply Prekopa-Leindler inequality for with

$$\lambda = \frac{1}{2}$$

$$h(x) = (2\pi)^{-n/2} e^{-\|x\|^2/2}$$

$$f(x) = \exp\left(\frac{d(x,A)^2}{4}\right) \cdot (2\pi)^{-n/2} e^{-\|x\|^2/2}$$

$$g(x) = 1_A \cdot (2\pi)^{-n/2} e^{-\|x\|^2/2}.$$

It only remains to check that they satisfy the assumption of Prekopa-Leindler:

$$h\left(\frac{x+y}{2}\right) \geq f(x)^{\frac{1}{2}} \cdot g(x)^{\frac{1}{2}}$$

for all x, y. This can be easily checked, where the parallelogram law is used. (without loss of generality, we may assume $y \in A$.) ∎

Remarks. The proof works for any log-concave density (not just Gaussian density).

13.4. Borell's inequality. Let us first recall that a measure $\mu : \mathcal{A} \to \mathbb{R}$ is *log-concave* if the function

$$A \in \mathcal{A} \mapsto \log \mu(A)$$

is concave. That is,

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^{\lambda} \mu(B)^{1-\lambda}$$

for all $A, B \in \mathcal{A}$ and for any $\lambda \in (0,1)$.

Example 13.6 (log-concave measures).
1. Lebesgue measure in $\mathbb{R}^n$. This follows from Brunn-Minkowski.
2. Normalized volume on any convex set. $\mu(A) = \mathrm{vol}(A \cap K)$ is log-concave. This follows from Prekopa-Leindler with

$$f = 1_{A\cap K}, g = 1_{B\cap K} \text{ and } h = 1_{\lambda A + (1-\lambda) B \cap K}.$$

3. Log-concave densities. By a result of C.Borell in 1975 ([**?**]), any measure with log-concave density is log-concave.

Theorem 13.7 (Borell's inequality). *Let μ be a log-concave probability measure on $\mathbb{R}^n$, and let $A \subset \mathbb{R}^n$ be symmetric and convex with $\mu(A) \geq \frac{2}{3}$. Then,*

$$\mu(tA) \geq 1 - e^{-ct}$$

for every $t > 0$.

We will prove a special case where the log-concave measure is the normalized volume on a convex body.

THEOREM 13.8 (Volume Distribution in Convex Sets). *Let $K \subset \mathbb{R}^n$ be a convex body (not necessarily symmetric). Consider the normalized volume measure*

$$\mu(A) = \frac{\operatorname{vol}(A \cap K)}{\operatorname{vol}(K)}.$$

Let A be the Euclidean ball with $\mu(A) = \frac{2}{3}$. Then

$$\mu\left((tA)^c\right) \leq e^{-ct}.$$

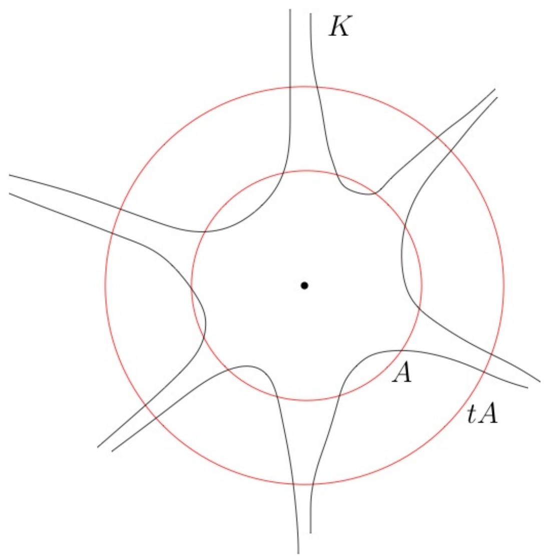

FIGURE 13. Volume Distribution of Convex Bodies

REMARKS. (1) One can also use this argument for $A =$ the slab between two hyperplanes with $\mu(A) = \frac{2}{3}$ and get the same result.
(2) Note that $\mu\left((tA)^c\right)$ describes the proportion of volume taken by the "tentacles" of the convex body.

PROOF OF THEOREM 13.8. Without loss of generality, assume that $t \geq 1$. We will find some convex combination of A and $(tA)^c$ that is disjoint from A. That is, we will find some $\lambda \in (0,1)$ such that

$$\lambda A + (1-\lambda)(tA)^c \subset A^c.$$

It can be easily checked that

$$\lambda = \frac{t-1}{t+1}$$

works. Then, by Brunn-Minkowski,

$$\frac{1}{3} \geq \mu(A^c) \geq \mu(\lambda A + (1-\lambda)(tA)^c) \geq \mu(A)^{\lambda} \mu((tA)^c)^{1-\lambda}.$$

Solve for $\mu((tA)^c)$, and we get the desired bound. ∎

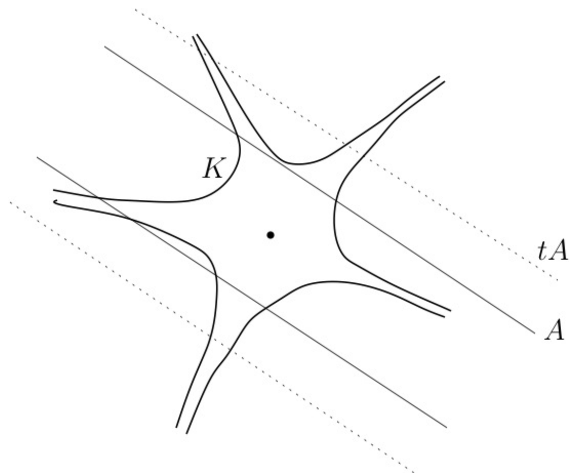

FIGURE 14. Volume Distribution between Hyperplanes

13.5. Urysohn's inequality. Recall that we discussed the two directions in considering Euclidean sections or projections of high dimensional convex bodies. One direction is Dvoretzky-type theorems, which are closely related to the mean width M^*. The other direction is VRT, which is closely related to the notion of volume ratio of a convex body. How do these two directions compare? Urysohn's inequality tells that the volume ratio is always dominated by the mean width.

THEOREM 13.9 (Urysohn's inequality; Chapter 1 in [**?**]). *Let $K \subset \mathbb{R}^n$ be a compact set. Then*

$$\left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(B_2^n)}\right)^{\frac{1}{n}} \leq \int_{S^{n-1}} \|x\|_{K^\circ} d\sigma(x).$$

As a preliminary to the proof, the following exercise gives alternative meaning of the mean width.

EXERCISE 18. Let $U \in O(n)$ be random. Then $\int_{O(n)} U(K) d\mu(U) = M^*(K) \cdot B_2^n$.

Now we are ready to prove Urysohn's inequality.

PROOF. First, it is not hard to see that Brunn-Minkowski inequality holds for more than two sets:

$$\mathrm{vol}\left(\sum_{i=1}^m \mathrm{vol}(K_i)\right)^{\frac{1}{n}} \geq \sum_{i=1}^m \mathrm{vol}(K_i)^{\frac{1}{n}}.$$

In the limit, we see that Brunn-Minkowski holds in integral form:

$$\mathrm{vol}\left(\int_{O(n)} U(K) d\mu(U)\right)^{\frac{1}{n}} \geq \int_{O(n)} \mathrm{vol}\left(U(K)\right)^{\frac{1}{n}} d\mu(U).$$

Then, by the exercise preceding this proof, we have

$$M^*(K) \cdot \mathrm{vol}(B_2^n)^{\frac{1}{n}} \geq \mathrm{vol}(K)^{\frac{1}{n}}.$$

■

13.6. Santalo inequality. In this application of Brunn-Minkowski inequality, we want to consider the "dual volume." We will discuss how the volume of a convex body is related to the volume of its polar body. We can easily picture that the polar set is "small" when the original set is "large" in terms of the volume, since they are sort of "inversely" related. One can see this through the fact that

$$(aK)^\circ = \frac{1}{a} K^\circ.$$

Hence, we might expect that the product of the two volumes behaves like a constant. We introduce the notion of *Mahler volume* for a better formulation of the problem.

Definition 13.1 (Malher volume). The *Mahler volume* of a convex body K is defined as

$$s(K) = \mathrm{vol}(K) \cdot \mathrm{vol}(K^\circ).$$

Exercise 19. Check that the Mahler volume is invariant under any linear transformation. For example, it is easy to see that it is invariant under scaling.

Hence, we may expect that there is an upper and a lower bound for $s(K)$, regardless of K. As we will see in this and next sections, finding these bounds are nontrivial. Different symmetrization technieques will be used in the two directions. The upper estimate uses the classic Steiner's symmetrization. For the lower estimate, things get more involved. We will use Milman's "isomorphic symmetrization."

Theorem 13.10 (Santalo inequality: upper estimate for Mahler volume). *Let K be a symmetric convex body in $\mathbb{R}^n$. Then $s(K) \leq s(B_2^n)$.*

Remarks.
(1) In other words, the Mahler volume is maximized by Euclidean balls. Moreover, by invariance under linear transformations, it is also maximized by any ellipsoid.
(2) $s(B_2^n) = \mathrm{vol}(B_2^n)^2 = [(2\pi e + o(1))n]^{-n}$.

We will present the proof due to Meyer and Pajor [**?**] using Steiner symmetrization. The idea is that these symmetrizations can only increase the Mahler volume, and they bring any symmetric convex body K closer and closer to the Euclidean ball.

Definition 13.2 (Steiner symmetrization). Let H be a hyperplane in $\mathbb{R}^n$ and $K \subset \mathbb{R}^n$. Then, the *Steiner symmetrization* of K with respect to H is

$$K_H = \left\{ \frac{x_1 - x_2}{2} : x_1, x_2 \in K, x_1 - x_2 \perp H \right\}.$$

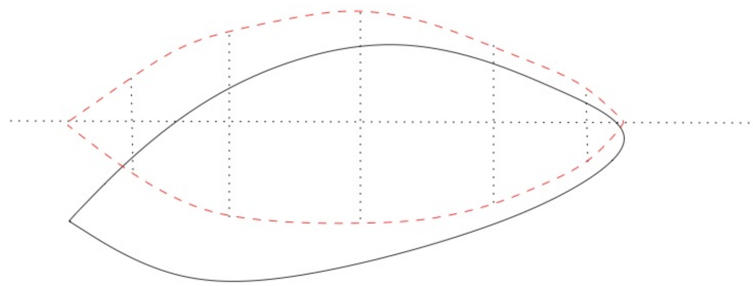

FIGURE 15. Steiner Symmetrization

Two main properties of Steiner symmetrizations are listed below. The reader is encouraged to check them as an exercise.

PROPOSITION 13.11 (Basic Properties of Steiner symmetrization).

(1) *Steiner symmetrization is volume-preserving.*
(2) *There exists a sequence of Steiner symmetrizations of any symmetric convex body K that converges to an Euclidean ball in Hausdorff metric (also in Banach-Mazur distance).*

Now, we will use Steiner symmetrization to prove Santalo inequality.

PROOF OF SANTALO INEQUALITY. It suffices to show that

$$\mathrm{vol}((K_H)^\circ) \geq \mathrm{vol}(K^\circ). \tag{17}$$

Indeed, if this holds, then

$$\mathrm{vol}(K_H) \cdot \mathrm{vol}((K_H)^\circ) \geq \mathrm{vol}(K) \cdot \mathrm{vol}(K^\circ),$$

as $\mathrm{vol}(K_H) = \mathrm{vol}(K)$. Now, by the second property of Steiner symmetrization, we get that in the limit,

$$\mathrm{vol}(B_2^n) \cdot \mathrm{vol}(B_2^n) \geq \mathrm{vol}(K) \cdot \mathrm{vol}(K^\circ),$$

which is exactly what we want to show.
Now, to prove (17), we may assume $H = \mathbb{R}^{n-1}$. Consider the slices

$$K(s) = \{x \in \mathbb{R}^{n-1} : (x, s) \in K\}.$$

It suffices to show that

$$\mathrm{vol}(K_H^\circ(s)) \geq \mathrm{vol}(K^\circ(s))$$

for all s, as

$$\mathrm{vol}(K^\circ) = \int \mathrm{vol}(K^\circ(y))dy.$$

It is easy to verify the following:

$$K^\circ = \{(y, s) : \langle x, y\rangle + ts \leq 1 \ \forall (x, t) \in K\}$$

$$K_H = \left\{(x, t) : t = \frac{t_1 - t_2}{2}, (x, t_i) \in K\right\}$$

$$(K_H)^\circ = \left\{ (y,s) : \langle x, y\rangle + \frac{t_1 - t_2}{2} \cdot s \le 1 \ \forall (x, t_i) \in K \right\}.$$

Then we claim that

$$K_H^\circ \supseteq \frac{K^\circ(s) + K^\circ(-s)}{2}.$$

In fact, let $y \in K^\circ(s)$ and $z \in K^\circ(-s)$. That is,

$$\langle x, y\rangle + t_1 s \le 1 \ \forall (x, t_1) \in K$$

$$\langle x, z\rangle - t_2 s \le 1 \ \forall (x, t_2) \in K.$$

Easy arithmetic operations give

$$\langle x, \frac{y+z}{2}\rangle + \frac{t_1 - t_2}{2} \cdot s \le 1.$$

This means

$$\frac{y+z}{2} \in K_H^\circ.$$

Hence, by Brunn-Minkowski inequality,

$$\begin{aligned} \mathrm{vol}(K_H^\circ) &\ge \mathrm{vol}\left(\frac{K^\circ(s) + K^\circ(-s)}{2} \right) \\ &\ge \sqrt{\mathrm{vol}(K^\circ(s)) \cdot \mathrm{vol}(K^\circ(-s))} \\ &= \mathrm{vol}(K^\circ(s)), \end{aligned}$$

as $K^\circ(-s) = -K^\circ(s)$ by symmetry. ■

14. Isomorphic Symmetrizations. Inverse Santalo Inequality

14.1. Inverse Santalo inequality. We have showed in Santalo inequality that Euclidean balls maximize the Mahler volume. A natural question is, which convex bodies minimize the Mahler volume?

CONJECTURE 14.1 (Mahler conjecture). *Cubes (or octahedra) are minimizers of the Mahler volume.*

Intuitively, this makes a lot of sense, as cubes and octahedra are the "farthest" from the Euclidean ball: they are the "pointiest" convex bodies. It is easy to check that

$$s(B_1^n) = s(B_\infty^n) = \mathrm{vol}(B_1^n) \cdot \mathrm{vol}(B_\infty^n) = (4e + o(1))^n n^{-n},$$

and

$$s(B_2^n) = \mathrm{vol}(B_2^n) \cdot \mathrm{vol}(B_2^n) = (2\pi e + o(1))^n n^{-n}.$$

Hence, the conjectured lower bound of Mahler volume matches the known upper bound: $s(B_1^n) = s(B_\infty^n) \sim c^n s(B_2^n)$ for some absolute constant $c > 0$. For more discussion, See Terry Tao's blog entry on this topic. Here, we will give a proof of a joint statement of both Santalo and inverse Santalo inequalities due to Bourgain and Milman. This proof provides some evidence that cubes and octahedra are good candidates for the minimizer of the Malher volume.

THEOREM 14.2 (Inverse Santalo inequality; [**?**]). *There exists an absolute constant $c \in (0,1)$ such that for every symmetric convex body K in $\mathbb{R}^n$,*

$$c^n \leq \frac{s(K)}{s(B_2^n)} \leq 1.$$

In other words,

$$c \leq \left(\frac{\mathrm{vol}(K) \cdot \mathrm{vol}(K^\circ)}{\mathrm{vol}(B_2^n)^2} \right)^{\frac{1}{n}} \leq 1.$$

The righthand side inequality is just Santalo inequality. Recall that in proving Santalo inequality, we used Steiner symmetrization, which is volume-preserving and can only increase the Mahler volume (i.e. increase the volume of the polar body). Moreover, there exists a sequence of Steiner symmetrizations that bring a symmetric convex body to a Euclidean ball. Hence, we cannot expect "exact" symmetrizations (meaning converging to the exact Euclidean ball) to get the inverse Santalo. Instead, we will have a sequence of symmetrizations that bring the convex body to C-isomorphic (instead of "isometric") to the Euclidean ball: we hope that in the end, we have

$$rB_2^n \subset K_t \subset C \cdot rB_2^n$$

for some $r > 0$. This way, we will also have

$$\frac{1}{rC} B_2^n \subset K_t^\circ \subset \frac{1}{r} B_2^n,$$

so that

$$\frac{1}{C^n} \mathrm{vol}(B_2^n)^2 \leq \mathrm{vol}(K_t)\, \mathrm{vol}(K_t^\circ) \leq C^n \,\mathrm{vol}(B_2^n)^2. \tag{18}$$

This is already very close to what we want: the only thing left to do is to make sure the Mahler volume is well controlled during symmetrization: $s(K_t)^{\frac{1}{n}} \asymp s(K)^{\frac{1}{n}}$ in the sense that there exist constants c_1 and c_2 such that

$$c_1^n s(K) \leq s(K_t) \leq c_2^n s(K). \tag{19}$$

In summary, we need to control two things in the proof of the inverse Santalo inequality:

(1) Distance from the Euclidean ball.
We will borrow the idea in the proof of the Quotient of Subspace Theorem. At each step, if

$$d(K_i, B_2^n) \lesssim M(K_{i-1}) M^*(K_{i-1}),$$

then the MM^*-estimate promises logarithmetically dropping distance.

(2) Volume control.
We will show

$$e^{-c_i n} s(K_i) \leq s(K_{i-1}) \leq e^{c_i n} s(K_i)$$

for some properly chosen constant c_i. After iterations, we will get

$$e^{-(\sum c_i)n} s(K) \leq s(K_t) \leq e^{(\sum c_i)n} s(K),$$

where our choice of c_i guarantees the convergence of the series.

Let us break down the proof into its main pieces, and then assemble them to form the complete proof. The first and most important piece is the symmetrization technique called "*isomorphic symmetrization*" or "*convex surgery*," as first called by Milman.

14.2. Isomorphic symmetrization. We transform K into

$$K_1 = \operatorname{conv}\left\{K \cap \lambda_{up} B_2^n, \frac{1}{\lambda_{down}} B_2^n\right\}.$$

Two operations are involved here to regularize K:

- Cut "tentacles" by intersecting K with a large Euclidean ball $\lambda_{up} B_2^n$, and
- Fill out the interior by taking convex hull with a small Euclidean ball $(1/\lambda_{down}) B_2^n$. See Figure 16.

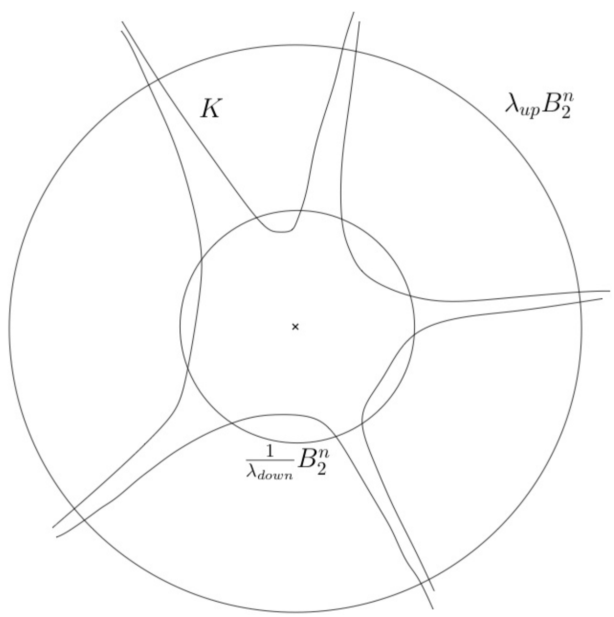

FIGURE 16. Isomorphic Symmetrization

Observe that

$$\frac{1}{\lambda_{down}} B_2^n \subset K_1 \subset \lambda_{up} B_2^n.$$

Hence,

$$d(K_1, B_2^n) \leq \lambda_{up} \cdot \lambda_{down} \tag{20}$$

Moreover,

$$\begin{aligned} K_1^\circ &= (K \cap \lambda_{up} B_2^n)^\circ \cap \lambda_{down} B_2^n \\ &= conv\Big\{K^\circ, \frac{1}{\lambda_{up}} B_2^n\Big\} \cap \lambda_{down} B_2^n \\ &\asymp \text{conv}\Big\{K^\circ \cap \lambda_{down} B_2^n, \frac{1}{\lambda_{up}} B_2^n\Big\}. \end{aligned}$$

For notation convenience, let $T = \text{conv}\Big\{K^\circ \cap \lambda_{down} B_2^n, \frac{1}{\lambda_{up}} B_2^n\Big\}$. We leave the last step above as an exercise.

Exercise 20. Show that there exists some $C \geq 1$ such that

$$\frac{1}{C} T \subset K_1^\circ \subset CT.$$

The second piece of the proof of the inverse Santalo inequality is how to control the volume of our convex bodies during symmetrizations. In particular, we want to control the volume of intersections and convex hulls of different convex bodies.

14.3. Volume control lemmas. First, we need to control the volume of intersections of two convex symmetric bodies.

Lemma 14.3 (Volume of intersections). *Let K, D be symmetric convex bodies in $\mathbb{R}^n$. Then*

$$\text{vol}(K) \leq N(K, D) \cdot \text{vol}(K \cap D).$$

Proof. Consider the covering of K by translates of D. Suppose that

$$K \subseteq \bigcup_{i=1}^{N} (x_i + D),$$

where $x_i \in K$. The reader should check that for any centrally symmetric convex bodies K and D, and for any $x \in \mathbb{R}^n$,

$$\text{vol}\big((x + D) \cap K\big) \leq \text{vol}(D \cap K).$$

Then, by symmetry and convexity, we have

$$\frac{(x + D) \cap K + (-x + D) \cap K}{2} \subseteq D \cap K.$$

Apply Brunn-Minkowski inequality, and we get

$$\text{vol}(D \cap K) \geq \text{vol}\big((x+D) \cap K\big)^{1/2} \cdot \text{vol}\big((-x+D) \cap K\big)^{1/2} = \text{vol}\big((x+D) \cap K\big).$$

Hence, $\text{vol}(K) \leq N(K, D)\, \text{vol}(K \cap D)$. ■

We also need to control the volume of convex hulls. We would like to have the following:

LEMMA 14.4 (Volume of Convex Hulls). *Let K, D be symmetric convex bodies in $\mathbb{R}^n$. Then*

$$\mathrm{vol}(\mathrm{conv}\{K, D\}) \leq N(D, K) \cdot \mathrm{vol}(K).$$

In Milman's paper [**?**], he proved a weaker result.

LEMMA 14.5 (Volume of convex hulls: weaker version). *Let K, D be convex sets in $\mathbb{R}^n$ such that $D \subset bK$ for some $b \geq 1$. Then,*

$$\mathrm{vol}(\mathrm{conv}\{K, D\}) \leq enb \cdot N(D, K) \cdot \mathrm{vol}(K).$$

We will prove an intermediate result, and leave it as an exercise to get to Milman's weaker version from there.

PROPOSITION 14.6. *Let $K, D \subset \mathbb{R}^n$ be convex bodies. Then, for any $\lambda \in (0, 1)$,*

$$\mathrm{vol}(\lambda K + (1-\lambda)D) \leq N(D, K) \cdot \mathrm{vol}(K).$$

PROOF. Cover D by $N = N(K, D)$ translates of K such that

$$D \subset \bigcup_{i=1}^{N} (x_i + K)$$

for some $\{x_i\}_{i=1}^N \subset D$. Multiply by $(1-\lambda)$ and add λK to get

$$\lambda K + (1-\lambda)D \subset \lambda K + \bigcup_{i=1}^{N} ((1-\lambda)x_i + (1-\lambda)K) = \bigcup_{i=1}^{N} (1-\lambda)x_i + K,$$

where the convexity of K is used in the last equality. Then, compare the volume. ∎

EXERCISE 21. Finish the proof for Lemma 14.5.
Hint: Note that

$$\begin{aligned} \mathrm{conv}\{K, D\} &= \bigcup_{\lambda \in [0,1]} \lambda K + (1-\lambda)D \\ &\subset \bigcup_{\lambda \in [0,1]} \bigcup_{i=1}^{N} (1-\lambda)x_i + K \\ &= \bigcup_{i=1}^{N} [0, x_i] + K. \end{aligned}$$

Then one can discretize the interval $[0, x_i]$ using $\sim n$ points and use union bound.

14.4. Proof of inverse Santalo inequality. Now we are in a position to put everything together to prove the inverse Santalo inequality (Theorem 14.2).

Proof of inverse Santalo inequality. We will follow two steps: isomorphic symmetrization and iterations.
Step 1: First isomorphic symmetrization.
Since the Mahler volume is invariant under linear transformations, we may assume that K is in an ℓ-position. So we have

$$M(K)M^*(K) \le C \log 2d(K, B_2^n) \le C \log n.$$

Choose

$$\lambda_{up} = M^*(K)a_1 \qquad \lambda_{down} = M(K)a_1,$$

where $a_1 \ge 1$ is to be determined later.
Consider the isomorphic symmetrization K_1 defined as

$$K_1 = \text{conv}\left\{K \cap \lambda_{up}B_2^n, \frac{1}{\lambda_{down}}B_2^n\right\}.$$

We claim that

$$e^{-Cn/a_1^2} \le \frac{\text{vol}(K_1)}{\text{vol}(K)} \le e^{Cn/a_1^2}.$$

In fact, by Lemma 14.3 and Sudakov inequality,

$$\begin{aligned}\text{vol}(K_1) &\ge \text{vol}\left(K \cap \lambda_{up}B_2^n\right)\\ &\ge \frac{\text{vol}(K)}{N\left(K, \lambda_{up}B_2^n\right)}\\ &\ge e^{-Cn/a_1^2} \cdot \text{vol}(K).\end{aligned}$$

This is the lower bound. For the upper bound, we use Lemma 14.4 and the dual Sudakov inequality.

$$\begin{aligned}\text{vol}(K_1) &\le \text{vol}\left(\text{conv}\left\{K, \frac{1}{\lambda_{down}}B_2^n\right\}\right)\\ &\le N\left(\frac{1}{\lambda_{down}}B_2^n, K\right) \cdot \text{vol}(K)\\ &\le e^{Cn/a_1^2} \cdot \text{vol}(K).\end{aligned}$$

The same computation can be done for K°, and we have

$$e^{-Cn/a_1^2} \le \frac{\text{vol}(K_1^\circ)}{\text{vol}(K^\circ)} \le e^{Cn/a_1^2}.$$

Hence,

$$e^{-Cn/a_1^2} \le \frac{s(K_1)}{s(K)} \le e^{Cn/a_1^2}.$$

At the same time, by (20) and MM^*-estimate,

$$\begin{aligned} d(K_1, B_2^n) &\leq \lambda_{up}\lambda_{down} \\ &\leq M^*(K)M(K) \cdot a_1^2 \\ &\leq Ca_1^2 \log 2d(K, B_2^n) \\ &\leq Ca_1^2 \log n. \end{aligned}$$

Choose $a_1 = \log n$, so that

$$d(K_1, B_2^n) \leq C(\log n)^3.$$

Step 2: Iteration
We first put K_1 in its ℓ-position. Then choose

$$\lambda_{up} = M^*(K_1)a_2 \qquad \lambda_{down} = M(K_1)a_2,$$

where $a_2 \geq 1$ is to be determined later.
Define K_2 in a similar way as above, and we get

$$e^{-Cn/a_2^2} \leq \frac{s(K_2)}{s(K_1)} \leq e^{Cn/a_2^2},$$

and

$$\begin{aligned} d(K_2, B_2^n) &\leq \lambda_{up}\lambda_{down} \leq M^*(K_1)M(K_1) \cdot a_2^2 \\ &\leq Ca_2^2 \log 2d(K_1, B_2^n) \leq Ca_2^2 \log\log n. \end{aligned}$$

Choose $a_2 = \log\log n$, so that

$$d(K_2, B_2^n) \leq C(\log\log n)^3.$$

After t iterations, we get

$$e^{-Cn\left(\frac{1}{a_1^2}+\cdots+\frac{1}{a_t^2}\right)} \leq \frac{s(K_t)}{s(K)} \leq e^{Cn\left(\frac{1}{a_1^2}+\cdots+\frac{1}{a_t^2}\right)}$$

and

$$d(K, B_2^n) \leq C\left(\log^{(t)} n\right)^3,$$

where $\log^{(t)} = \log\log\cdots\log$ denotes n iterations of log. Choose t to be the smallest integer such that

$$\log^{(t)} n < 2.$$

Then

$$d(K_t, B_2^n) \leq C_1$$

for some absolute constant $C_1 > 0$. Moreover, since with our choice of the a_i's, the series $\sum(1/a_i^2)$ is uniformly convergent, we obtain that for some absolute constant $C_2 > 0$,

$$e^{-C_2 n} \leq \frac{s(K_t)}{s(K)} \leq e^{C_2 n}.$$

Then, by (18) and (19), the proof is complete. ∎

An immediate corollary says that the Mahler volume of any symmetric convex bodies in $\mathbb{R}^n$ are equivalent.

Corollary 14.7 (Malher volume of arbitrary symmetric convex bodies). *For any symmetric convex bodies $K, L \subset \mathbb{R}^n$, there exists an absolute constant $c \in (0,1)$ such that*

$$c^n \le \frac{s(K)}{s(L)} \le \frac{1}{c^n}.$$

15. Applications of Isomorphic Symmetrizations

15.1. Milman's ellipsoids. The method of isomorphic symmetrization turns out to be quite flexible. Instead of controlling $\mathrm{vol}(K)$ and $\mathrm{vol}(K^\circ)$ throughout symmetrizations, one can control $\mathrm{vol}(K+L)$ for an arbitrary symmetric convex body L. This idea leads to the existence of so-called M-ellipsoids.

Theorem 15.1 (M-ellipsoids). *For every symmetric convex $K \subset \mathbb{R}^n$, there exists an ellipsoid $\mathcal{E}_K$ called an M-*ellipsoid *with the following properties:*

(1) $\mathrm{vol}(\mathcal{E}_K) = \mathrm{vol}(K)$;
(2) *There exists an absolute constant $C \ge 1$ such that*

$$\frac{1}{C^n}\,\mathrm{vol}(\mathcal{E}_K + L) \le \mathrm{vol}(K+L) \le C^n\,\mathrm{vol}(\mathcal{E}_K + L)$$

for every symmetric convex body $L \subset \mathbb{R}^n$.

Remarks. The existence of M-ellipsoids means that a symmetric convex body "behaves" like an ellipsoid.

The proof is similar to the proof of the inverse Santalo inequality in the previous section. So we will only sketch the proof and leave the details as an exercise. Before starting the proof, we will have to modify our volume control lemmas. We encourage the reader to verify them in a similar way to their original versions.

Lemma 15.2 (Modified volume of intersection).

$$\mathrm{vol}(K+L) \le N(K,D)\,\mathrm{vol}(K \cap D + L).$$

Lemma 15.3 (Modified Volume of Convex Hull).

$$\mathrm{vol}(\mathrm{conv}\{K,D\} + L) \le 2enb \cdot N(K,D) \cdot \mathrm{vol}(K+L).$$

Proof of Theorem 15.1. Define the isomorphic symmetrizations K_i the same way as before. At each step, we have

$$e^{-Cn/a_i^2} \le \frac{\mathrm{vol}(K_i+L)}{\mathrm{vol}(K_{i-1}+L)} \le e^{Cn/a_i^2} \tag{21}$$

and

$$d(K_t, B_2^n) \le C a_i^2 \log d(K_{i-1}, B-2^n),$$

where $a_i = \log^{(i)} n$. Hence, we conclude that

$$e^{-C_1 n} \le \frac{\mathrm{vol}(K_t + L)}{\mathrm{vol}(K + L)} \le e^{C_1 n} \qquad \forall\ L,$$

while $d(K_t, B_2^n) \le C_2$, which means that K_t is C-isomorphic to some ellipsoid $\mathcal{E}$. Therefore,

$$\mathrm{vol}(K)^{\frac{1}{n}} \sim \mathrm{vol}(K_t)^{\frac{1}{n}} \sim \mathrm{vol}(\mathcal{E})^{\frac{1}{n}}, \tag{22}$$

where the first $\sim$ follows from (21) and the second $\sim$ follows from the C-isomorphism. Now, let $\mathcal{E}_K = \rho\mathcal{E}$ such that $\mathrm{vol}(K) = \mathrm{vol}(\rho\mathcal{E})$. By (22), we see that

$$\rho \sim \text{ constant.}$$

Since K_t is still $O(1)$-isomorphic to $\mathcal{E}_K$, we can replace K_t by $\mathcal{E}_K$ in (21). ■

Remarks. (1) *Uniqueness of M-ellipsoids.* M-ellipsoid is not unique, as the constant C is not specified. We leave it an exercise to check that one can dilate one axis of an M-ellipsoid by 2, and it still remains an M-ellipsoid.

(2) *Duality.* Similar with the proof of the inverse Santalo inequality, one can run isomorphic symmetrizations for K and K° in parallel. This way, one gets

$$\mathcal{E}_{K^\circ} = \rho \cdot (\mathcal{E}_K)^\circ$$

in the end, where $\rho \sim$ constant. Note that one can define M-ellipsoids in an isomorphic way as

$$\mathrm{vol}(\mathcal{E}_K) \sim \mathrm{vol}(K),$$

that is,

$$\frac{1}{C^n}\,\mathrm{vol}(K) \le \mathrm{vol}(\mathcal{E}_K) \le C^n\,\mathrm{vol}(K),$$

rather than exact equality. Adopting this definition, we have

$$\mathcal{E}_{K^\circ} = (\mathcal{E}_K)^\circ.$$

(3) *M-position.* By Theorem 15.1, we can define a position of a convex body called an *M-position*.

Definition 15.1 (M-position). Let K be a convex body in $\mathbb{R}^n$. Then, K is said to be in an *M-position* if its M-ellipsoid is an Eucdliean ball.

15.2. Inverse Brunn-Minkowski inequality. Without putting on extra conditions on the convex bodies, the reader is encouraged to find examples that fail the inverse Brunn-Minkowski inequality (this is not hard). However, if we assume in addition that the convex bodies are in M-positions, then the inverse Brunn-Minkowski is a quick consequence of Theorem 15.1.

Corollary 15.4 (Inverse Brunn-Minkowski inequality). *Let $K, L \subset \mathbb{R}^n$ be symmetric convex bodies whose M-ellipsoids are homothetic. Then*

$$\operatorname{vol}(K+L)^{\frac{1}{n}} \leq C\left(\operatorname{vol}(K)^{\frac{1}{n}} + \operatorname{vol}(L)^{\frac{1}{n}}\right).$$

Proof. We will use Theorem 15.1 twice.

$$\begin{aligned}
\operatorname{vol}(K+L)^{\frac{1}{n}} &\leq C \operatorname{vol}(\mathcal{E}_K + L)^{\frac{1}{n}} \\
&\leq C^2 \operatorname{vol}(\mathcal{E}_K + \mathcal{E}_L)^{\frac{1}{n}} \\
&= C^2\left(\operatorname{vol}(\mathcal{E}_K)^{\frac{1}{n}} + \operatorname{vol}(\mathcal{E}_L))^{\frac{1}{n}}\right) \qquad \text{(by homothety).}
\end{aligned}$$

■

Remarks. By this corollary, we see that here is always a position of K such that the reverse Brunn-Minkowski inequality holds.

The following is a consequence of inverse Brunn-Minkowski inequality.

Corollary 15.5. *Let $K, L \subset \mathbb{R}^n$ be symmetric convex bodies with homothetic M-ellipsoids. Then*

$$\operatorname{vol}(K \cap L)^{\frac{1}{n}} \geq c \cdot \min\left\{\operatorname{vol}(K)^{\frac{1}{n}}, \operatorname{vol}(L)^{\frac{1}{n}}\right\}.$$

Proof. Let $v(K)$ denote the volume ratio $\left(\operatorname{vol}(K)/\operatorname{vol}(B_2^n)\right)^{1/n}$. Then, by Santalo and inverse Santalo inequalities, we have $v(K^\circ) \sim 1/v(K)$. Then,

$$\begin{aligned}
\frac{1}{v(K \cap L)} &\sim v((K \cap L)^\circ) \\
&\sim v(K^\circ + L^\circ) \quad (\text{as } (K \cap L)^\circ = \mathit{conv}\{K^\circ, L^\circ\} \asymp K^\circ + L^\circ) \\
&\sim v(K^\circ) + v(L^\circ) \quad \text{(by BM and reverse BM)} \\
&\sim \frac{1}{v(K)} + \frac{1}{v(L)} \quad \text{(by Santalo and reverse Santalo).}
\end{aligned}$$

This completes the proof. ■

Another consequence of inverse Brunn-Minkowski inequality is that the covering number and the volume ratio of two convex bodies with homothetic M-ellipsoids are equivalent.

Corollary 15.6 (Covering number and volume ratio). *Let $K, L \subset \mathbb{R}^n$ be symmetric convex bodies with homothetic M-ellipsoids. Then*

$$\frac{\operatorname{vol}(K)}{\operatorname{vol}(L)} \leq N(K, L) \leq C^n \frac{\operatorname{vol}(K)}{\operatorname{vol}(L)},$$

provided that $\frac{\operatorname{vol}(K)}{\operatorname{vol}(L)} \geq 1$.

PROOF.

$$\begin{aligned} N(K,L)^{\frac{1}{n}} &\le \frac{\mathrm{vol}(K+\frac{1}{2}L)}{\mathrm{vol}(\frac{1}{2}L)} \quad \text{(by covering number estimate)} \\ &\le C\cdot \frac{\mathrm{vol}(K)^{\frac{1}{n}}+\mathrm{vol}(\frac{1}{2}L)^{\frac{1}{n}}}{\mathrm{vol}(\frac{1}{2}L)^{\frac{1}{n}}} \quad \text{(by reverse Brunn-Minkowski)} \\ &\le C\left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(L)}\right)^{\frac{1}{n}}. \end{aligned}$$

■

When we consider the covering of a symmetric convex body by its M-ellipsoid (or vice versa), we have the following result as a direct consequence of the above corollary.

COROLLARY 15.7 (Covering number by M-ellipsoids). *There exist absolute constants $c_1, c_2 > 0$ such that for every symmetric convex body $K \subset \mathbb{R}^n$,*

$$N(K,\mathcal{E}_K) \le e^{c_1 n} \qquad N(\mathcal{E}_K, K) \le e^{c_2 n}.$$

15.3. Duality of entropy: on the exponential scale. In this subsection, we will present a weak result on the duality of metric entropy. First, let us see what we have on the duality of volume ratios.

COROLLARY 15.8 (Volume ratio duality). *Let K and L be symmetric and convex bodies in $\mathbb{R}^n$. Consider the volume ratio defined by*

$$v(K,L) = \left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(L)}\right)^{\frac{1}{n}}.$$

Then, there exists an absolute constant $C \ge 1$ such that

$$\frac{1}{C}v(L^\circ,K^\circ) \le v(K,L) \le Cv(L^\circ,K^\circ).$$

SKETCH OF PROOF. Note that by Theorem 14.2,

$$\frac{v(K,L)}{v(L^\circ,K^\circ)} = (s(K)s(L))^{\frac{1}{n}} \sim \text{constant}.$$

■

Now, we present a result on the duality of metric entropy due to Konig and Milman.

COROLLARY 15.9 (Entropy duality; [?]). *Let $K, L \subset \mathbb{R}^n$ be symmetric and convex. Then there exists an absolute constant $C > 1$ such that*

$$\frac{1}{C^n}N(L^\circ,K^\circ) \le N(K,L) \le C^n N(L^\circ,K^\circ).$$

PROOF. By Lemma 14.3 and the inverse Santalo inequality,

$$N(K,L) \ge \frac{\mathrm{vol}(K)}{\mathrm{vol}(K\cap L)} \ge c^n \frac{\mathrm{vol}(K\cap L)^\circ}{\mathrm{vol}(K^\circ)}.$$

Note that

$$(K \cap L)^\circ = \mathrm{conv}\{K^\circ, L^\circ\} \supset \frac{K^\circ + L^\circ}{2}.$$

Then,

$$N(K, L) \geq c^n \frac{\mathrm{vol}(K^\circ + L^\circ)}{\mathrm{vol}(K^\circ/2} \geq c^n \frac{\mathrm{vol}(K^\circ/2 + L^\circ)}{\mathrm{vol}(K^\circ/2}) \geq c^n N(L^\circ.K^\circ),$$

where the last step follows from the covering number estimate

$$N(K, L) \leq \frac{\mathrm{vol}\left(K + \frac{1}{2}L\right)}{\mathrm{vol}\left(\frac{1}{2}L\right)}.$$

∎

Remarks. This corollary solves a weak form of the duality conjecture

$$N(L^\circ, CK^\circ)^c \leq N(K, L) \leq N(L^\circ, cK^\circ)^C$$

when the covering number $N(K, L) \sim e^{cn}$, where an exponential coefficient does not ruin the order.

15.4. An alternative proof of Quotient of Subspace Theorem. Our last application is a quotient of subspace theorem where the subspaces are random. Recall that the QS-theorem we proved in only gives existence of such subspaces. In this subsection, we will see that it actually holds with high probability under some M-position assumptions.

Corollary 15.10. *(Random Quotient of Subspace theorem) Let $K \subset \mathbb{R}^n$ be a symmetric convex body in an M-position, let $\delta \in (0, 1)$, and let E be a random $(1-\delta)n$-dimensional subspace and $F \subset E$ a random $k = (1-\delta)^2 n$-dimensional subspace. Then with probability at least $1 - e^{-ck}$,*

$$d\left(P_F(E \cap K), B_2^k\right) \leq C(\delta),$$

where $C(\delta)$ is a constant that depends only on δ.

Proof. As K is in an M-position, we may assume that its M-ellipsoid $\mathcal{E}_K = B_2^n$, and $\mathcal{E}_{K^\circ} = B_2^n$. By Corollary 15.7,

$$N(K, B_2^n) \leq e^{c_1 n} \qquad N(K^\circ, B_2^n) \leq e^{c_2 n}.$$

Moreover, by Theorem 7.6, a random subspace E of dimension $(1 - \delta)n$ satisfies that, with high probability,

$$\mathrm{diam}(K \cap E) \leq C(\delta).$$

That is, $K \cap E \subset C(\delta) \cdot B_2^n \cap E$. By duality, we get

$$P_E(K^\circ) \supset \frac{1}{C(\delta)} \cdot P_E(B_2^n).$$

Hence, we found a Euclidean ball inside a projection of K°. Clearly,

$$N\Big(P_E(K^\circ), P_E(B_2^n)\Big) \leq N(K^\circ, B_2^n) \leq e^{c_2 n}.$$

Apply entropy theorem again, for $P_E(K^\circ)$ and a random $(1-\delta)^2 n$-dimensional subspace F,

$$\mathrm{diam}(P_E(K^\circ) \cap F) \leq C(\delta).$$

Consequently,

$$\frac{1}{C(\delta}) \cdot B_2^n \cap F \subset P_E(K^\circ) \cap F \subset C(\delta) \cdot B_2^n \cap F.$$

Hence, by duality,

$$d(P_F(K \cap E), B_2^k) = d(P_E(K^\circ) \cap F, B_2^k) \leq C(\delta)^2.$$

■