Flavors of Geometry
MSRI Publications
Volume **31**, 1997

An Elementary Introduction to Modern Convex Geometry

KEITH BALL

Contents

Preface

These notes are based, somewhat loosely, on three series of lectures given by myself, J. Lindenstrauss and G. Schechtman, during the Introductory Workshop in Convex Geometry held at the Mathematical Sciences Research Institute in Berkeley, early in 1996. A fourth series was given by B. Bollobás, on rapid mixing and random volume algorithms; they are found elsewhere in this book.

The material discussed in these notes is not, for the most part, very new, but the presentation has been strongly influenced by recent developments: among other things, it has been possible to simplify many of the arguments in the light of later discoveries. Instead of giving a comprehensive description of the state of the art, I have tried to describe two or three of the more important ideas that have shaped the modern view of convex geometry, and to make them as accessible as possible to a broad audience. In most places, I have adopted an informal style that I hope retains some of the spontaneity of the original lectures. Needless to say, my fellow lecturers cannot be held responsible for any shortcomings of this presentation.

I should mention that there are large areas of research that fall under the very general name of convex geometry, but that will barely be touched upon in these notes. The most obvious such area is the classical or "Brunn–Minkowski" theory, which is well covered in [Schneider 1993]. Another noticeable omission is the combinatorial theory of polytopes: a standard reference here is [Brøndsted 1983].

Lecture 1. Basic Notions

The topic of these notes is convex geometry. The objects of study are convex bodies: compact, convex subsets of Euclidean spaces, that have nonempty interior. Convex sets occur naturally in many areas of mathematics: linear programming, probability theory, functional analysis, partial differential equations, information theory, and the geometry of numbers, to name a few.

Although convexity is a simple property to formulate, convex bodies possess a surprisingly rich structure. There are several themes running through these notes: perhaps the most obvious one can be summed up in the sentence: "All convex bodies behave a bit like Euclidean balls." Before we look at some ways in which this is true it is a good idea to point out ways in which it definitely is not. This lecture will be devoted to the introduction of a few basic examples that we need to keep at the backs of our minds, and one or two well known principles.

The only notational conventions that are worth specifying at this point are the following. We will use $|\cdot|$ to denote the standard Euclidean norm on $\mathbb{R}^n$. For a body K, $\mathrm{vol}(K)$ will mean the volume measure of the appropriate dimension.

The most fundamental principle in convexity is the *Hahn–Banach separation theorem*, which guarantees that each convex body is an intersection of half-spaces, and that at each point of the boundary of a convex body, there is at least one supporting hyperplane. More generally, if K and L are disjoint, compact, convex subsets of $\mathbb{R}^n$, then there is a linear functional $\phi : \mathbb{R}^n \to \mathbb{R}$ for which $\phi(x) < \phi(y)$ whenever $x \in K$ and $y \in L$.

The simplest example of a convex body in $\mathbb{R}^n$ is the cube, $[-1,1]^n$. This does not look much like the Euclidean ball. The largest ball inside the cube has radius 1, while the smallest ball containing it has radius $\sqrt{n}$, since the corners of the cube are this far from the origin. So, as the dimension grows, the cube resembles a ball less and less.

The second example to which we shall refer is the n-dimensional regular solid simplex: the convex hull of $n+1$ equally spaced points. For this body, the ratio of the radii of inscribed and circumscribed balls is n: even worse than for the cube. The two-dimensional case is shown in Figure 1. In Lecture 3 we shall see

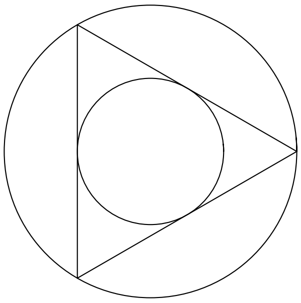

Figure 1. Inscribed and circumscribed spheres for an n-simplex.

that these ratios are extremal in a certain well-defined sense.

Solid simplices are particular examples of cones. By a *cone* in $\mathbb{R}^n$ we just mean the convex hull of a single point and some convex body of dimension $n-1$ (Figure 2). In $\mathbb{R}^n$, the volume of a cone of height h over a base of $(n-1)$-dimensional volume B is Bh/n.

The third example, which we shall investigate more closely in Lecture 4, is the n-dimensional "octahedron", or *cross-polytope:* the convex hull of the $2n$ points $(\pm 1, 0, 0, \ldots, 0)$, $(0, \pm 1, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, \pm 1)$. Since this is the unit ball of the ℓ_1 norm on $\mathbb{R}^n$, we shall denote it B_1^n. The circumscribing sphere of B_1^n has radius 1, the inscribed sphere has radius $1/\sqrt{n}$; so, as for the cube, the ratio is $\sqrt{n}$: see Figure 3, left.

B_1^n is made up of 2^n pieces similar to the piece whose points have nonnegative coordinates, illustrated in Figure 3, right. This piece is a cone of height 1 over a base, which is the analogous piece in $\mathbb{R}^{n-1}$. By induction, its volume is

$$\frac{1}{n} \cdot \frac{1}{n-1} \cdot \cdots \cdot \frac{1}{2} \cdot 1 = \frac{1}{n!},$$

and hence the volume of B_1^n is $2^n/n!$.

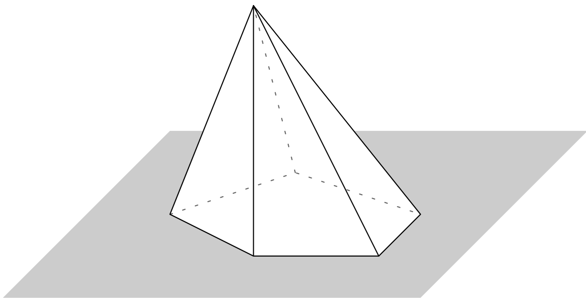

Figure 2. A cone.

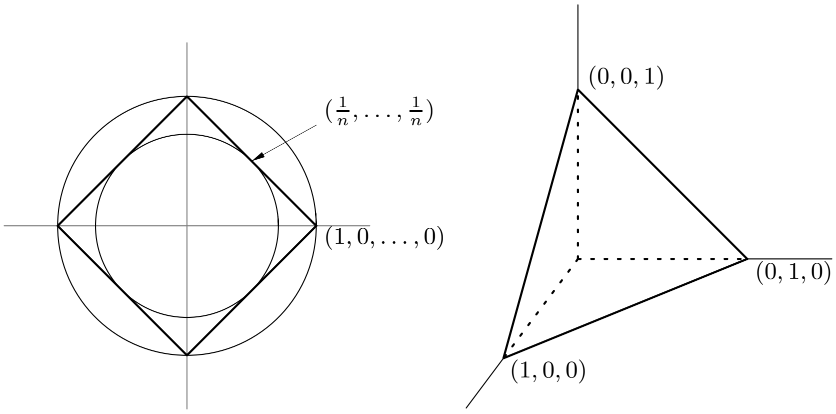

Figure 3. The cross-polytope (left) and one orthant thereof (right).

The final example is the Euclidean ball itself,

$$B_2^n = \left\{ x \in \mathbb{R}^n : \sum_1^n x_i^2 \leq 1 \right\}.$$

We shall need to know the volume of the ball: call it v_n. We can calculate the surface "area" of B_2^n very easily in terms of v_n: the argument goes back to the ancients. We think of the ball as being built of thin cones of height 1: see Figure 4, left. Since the volume of each of these cones is $1/n$ times its base area, the surface of the ball has area nv_n. The sphere of radius 1, which is the surface of the ball, we shall denote S^{n-1}.

To calculate v_n, we use integration in spherical polar coordinates. To specify a point x we use two coordinates: r, its distance from 0, and θ, a point on the sphere, which specifies the direction of x. The point θ plays the role of $n-1$ real coordinates. Clearly, in this representation, $x = r\theta$: see Figure 4, right. We can

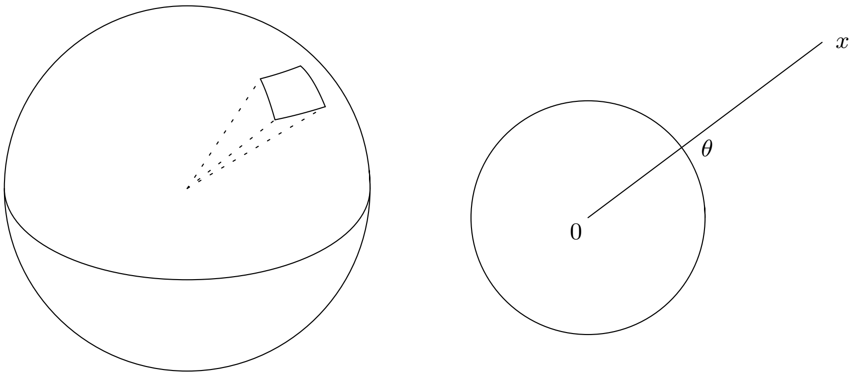

Figure 4. Computing the volume of the Euclidean ball.

write the integral of a function on $\mathbb{R}^n$ as

$$\int_{\mathbb{R}^n} f = \int_{r=0}^{\infty} \int_{S^{n-1}} f(r\theta) \text{ ``}d\theta\text{'' } r^{n-1}\, dr. \tag{1.1}$$

The factor r^{n-1} appears because the sphere of radius r has area r^{n-1} times that of S^{n-1}. The notation "$d\theta$" stands for "area" measure on the sphere: its total mass is the surface area nv_n. The most important feature of this measure is its rotational invariance: if A is a subset of the sphere and U is an orthogonal transformation of $\mathbb{R}^n$, then UA has the same measure as A. Throughout these lectures we shall normalise integrals like that in (1.1) by pulling out the factor nv_n, and write

$$\int_{\mathbb{R}^n} f = nv_n \int_0^{\infty} \int_{S^{n-1}} f(r\theta) r^{n-1}\, d\sigma(\theta)\, dr$$

where $\sigma = \sigma_{n-1}$ is the rotation-invariant measure on S^{n-1} of total mass 1. To find v_n, we integrate the function

$$x \mapsto \exp\left(-\tfrac{1}{2}\sum_1^n x_i^2\right)$$

both ways. This function is at once invariant under rotations and a product of functions depending upon separate coordinates; this is what makes the method work. The integral is

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} \prod_1^n e^{-x_i^2/2}\, dx = \prod_1^n \left(\int_{-\infty}^{\infty} e^{-x_i^2/2}\, dx_i\right) = \left(\sqrt{2\pi}\right)^n.$$

But this equals

$$nv_n \int_0^{\infty} \int_{S^{n-1}} e^{-r^2/2} r^{n-1}\, d\sigma\, dr = nv_n \int_0^{\infty} e^{-r^2/2} r^{n-1}\, dr = v_n 2^{n/2} \Gamma\left(\frac{n}{2}+1\right).$$

Hence

$$v_n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}.$$

This is extremely small if n is large. From Stirling's formula we know that

$$\Gamma\left(\frac{n}{2}+1\right) \sim \sqrt{2\pi}\, e^{-n/2} \left(\frac{n}{2}\right)^{(n+1)/2},$$

so that v_n is roughly

$$\left(\sqrt{\frac{2\pi e}{n}}\right)^n.$$

To put it another way, the Euclidean ball of *volume* 1 has *radius* about

$$\sqrt{\frac{n}{2\pi e}},$$

which is pretty big.

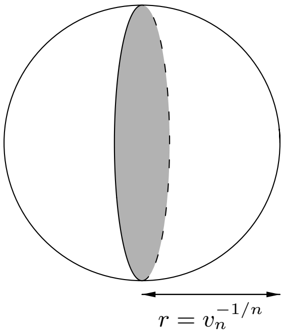

Figure 5. Comparing the volume of a ball with that of its central slice.

This rather surprising property of the ball in high-dimensional spaces is perhaps the first hint that our intuition might lead us astray. The next hint is provided by an answer to the following rather vague question: how is the mass of the ball distributed? To begin with, let's estimate the $(n-1)$-dimensional volume of a slice through the centre of the ball of volume 1. The ball has radius

$$r = v_n^{-1/n}$$

(Figure 5). The slice is an $(n-1)$-dimensional ball of this radius, so its volume is

$$v_{n-1} r^{n-1} = v_{n-1} \left(\frac{1}{v_n} \right)^{(n-1)/n}.$$

By Stirling's formula again, we find that the slice has volume about $\sqrt{e}$ when n is large. What are the $(n-1)$-dimensional volumes of parallel slices? The slice at distance x from the centre is an $(n-1)$-dimensional ball whose radius is $\sqrt{r^2 - x^2}$ (whereas the central slice had radius r), so the volume of the smaller slice is about

$$\sqrt{e} \left(\frac{\sqrt{r^2 - x^2}}{r} \right)^{n-1} = \sqrt{e} \left(1 - \frac{x^2}{r^2} \right)^{(n-1)/2}.$$

Since r is roughly $\sqrt{n/(2\pi e)}$, this is about

$$\sqrt{e} \left(1 - \frac{2\pi e x^2}{n} \right)^{(n-1)/2} \approx \sqrt{e} \exp(-\pi e x^2).$$

Thus, if we project the mass distribution of the ball of volume 1 onto a single direction, we get a distribution that is approximately Gaussian (normal) with variance $1/(2\pi e)$. What is remarkable about this is that the variance does not depend upon n. Despite the fact that the radius of the ball of volume 1 grows like $\sqrt{n/(2\pi e)}$, almost all of this volume stays within a slab of fixed width: for example, about 96% of the volume lies in the slab

$$\{x \in \mathbb{R}^n : -\tfrac{1}{2} \le x_1 \le \tfrac{1}{2}\}.$$

See Figure 6.

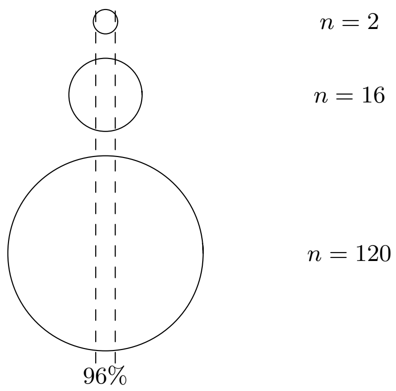

Figure 6. Balls in various dimensions, and the slab that contains about 96% of each of them.

So the volume of the ball concentrates close to *any* subspace of dimension $n-1$. This would seem to suggest that the volume concentrates near the centre of the ball, where the subspaces all meet. But, on the contrary, it is easy to see that, if n is large, most of the volume of the ball lies near its surface. In objects of high dimension, measure tends to concentrate in places that our low-dimensional intuition considers small. A considerable extension of this curious phenomenon will be exploited in Lectures 8 and 9.

To finish this lecture, let's write down a formula for the volume of a general body in spherical polar coordinates. Let K be such a body with 0 in its interior, and for each direction $\theta \in S^{n-1}$ let $r(\theta)$ be the radius of K in this direction. Then the volume of K is

$$nv_n \int_{S^{n-1}} \int_0^{r(\theta)} s^{n-1}\, ds\, d\sigma = v_n \int_{S^{n-1}} r(\theta)^n \, d\sigma(\theta).$$

This tells us a bit about particular bodies. For example, if K is the cube $[-1,1]^n$, whose volume is 2^n, the radius satisfies

$$\int_{S^{n-1}} r(\theta)^n = \frac{2^n}{v_n} \approx \left(\sqrt{\frac{2n}{\pi e}}\right)^n.$$

So the "average" radius of the cube is about

$$\sqrt{\frac{2n}{\pi e}}.$$

This indicates that the volume of the cube tends to lie in its corners, where the radius is close to $\sqrt{n}$, not in the middle of its facets, where the radius is close to 1. In Lecture 4 we shall see that the reverse happens for B_1^n and that this has a surprising consequence.

If K is (*centrally*) *symmetric*, that is, if $-x \in K$ whenever $x \in K$, then K is the unit ball of some norm $\|\cdot\|_K$ on $\mathbb{R}^n$:

$$K = \{x : \|x\|_K \le 1\}.$$

This was already mentioned for the octahedron, which is the unit ball of the ℓ_1 norm

$$\|x\| = \sum_1^n |x_i|.$$

The norm and radius are easily seen to be related by

$$r(\theta) = \frac{1}{\|\theta\|}, \quad \text{for } \theta \in S^{n-1},$$

since $r(\theta)$ is the largest number r for which $r\theta \in K$. Thus, for a general symmetric body K with associated norm $\|\cdot\|$, we have this formula for the volume:

$$\mathrm{vol}(K) = v_n \int_{S^{n-1}} \|\theta\|^{-n}\, d\sigma(\theta).$$

Lecture 2. Spherical Sections of the Cube

In the first lecture it was explained that the cube is rather unlike a Euclidean ball in $\mathbb{R}^n$: the cube $[-1,1]^n$ includes a ball of radius 1 and no more, and is included in a ball of radius $\sqrt{n}$ and no less. The cube is a bad approximation to the Euclidean ball. In this lecture we shall take this point a bit further. A body like the cube, which is bounded by a finite number of flat facets, is called a *polytope*. Among symmetric polytopes, the cube has the fewest possible facets, namely $2n$. The question we shall address here is this:

> *If K is a polytope in $\mathbb{R}^n$ with m facets, how well can K approximate the Euclidean ball?*

Let's begin by clarifying the notion of approximation. To simplify matters we shall only consider symmetric bodies. By analogy with the remarks above, we could define the distance between two convex bodies K and L to be the smallest d for which there is a scaled copy of L inside K and another copy of L, d times as large, containing K. However, for most purposes, it will be more convenient to have an affine-invariant notion of distance: for example we want to regard all parallelograms as the same. Therefore:

Definition. The distance $d(K, L)$ between symmetric convex bodies K and L is the least positive d for which there is a linear image $\tilde{L}$ of L such that $\tilde{L} \subset K \subset d\tilde{L}$. (See Figure 7.)

Note that this distance is multiplicative, not additive: in order to get a metric (on the set of linear equivalence classes of symmetric convex bodies) we would need to take $\log d$ instead of d. In particular, if K and L are identical then $d(K, L) = 1$.

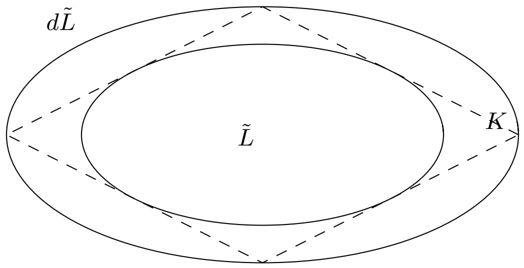

Figure 7. Defining the distance between K and L.

Our observations of the last lecture show that the distance between the cube and the Euclidean ball in $\mathbb{R}^n$ is *at most* $\sqrt{n}$. It is intuitively clear that it really is $\sqrt{n}$, i.e., that we cannot find a linear image of the ball that sandwiches the cube any better than the obvious one. A formal proof will be immediate after the next lecture.

The main result of this lecture will imply that, if a polytope is to have small distance from the Euclidean ball, it must have very many facets: exponentially many in the dimension n.

THEOREM 2.1. *Let K be a (symmetric) polytope in $\mathbb{R}^n$ with $d(K, B_2^n) = d$. Then K has at least $e^{n/(2d^2)}$ facets. On the other hand, for each n, there is a polytope with 4^n facets whose distance from the ball is at most* 2.

The arguments in this lecture, including the result just stated, go back to the early days of packing and covering problems. A classical reference for the subject is [Rogers 1964].

Before we embark upon a proof of Theorem 2.1, let's look at a reformulation that will motivate several more sophisticated results later on. A symmetric convex body in $\mathbb{R}^n$ with m pairs of facets can be realised as an n-dimensional slice (through the centre) of the cube in $\mathbb{R}^m$. This is because such a body is the intersection of m slabs in $\mathbb{R}^n$, each of the form $\{x : |\langle x, v\rangle| \le 1\}$, for some nonzero vector v in $\mathbb{R}^n$. This is shown in Figure 8.

Thus K is the set $\{x : |\langle x, v_i\rangle| \le 1 \text{ for } 1 \le i \le m\}$, for some sequence $(v_i)_1^m$ of vectors in $\mathbb{R}^n$. The linear map

$$T : x \mapsto (\langle x, v_1\rangle, \ldots, \langle x, v_m\rangle)$$

embeds $\mathbb{R}^n$ as a subspace H of $\mathbb{R}^m$, and the intersection of H with the cube $[-1, 1]^m$ is the set of points y in H for which $|y_i| \le 1$ for each coordinate i. So this intersection is the image of K under T. Conversely, any n-dimensional slice of $[-1, 1]^m$ is a body with at most m pairs of faces. Thus, the result we are aiming at can be formulated as follows:

> *The cube in $\mathbb{R}^m$ has almost spherical sections whose dimension n is roughly* $\log m$ *and not more.*

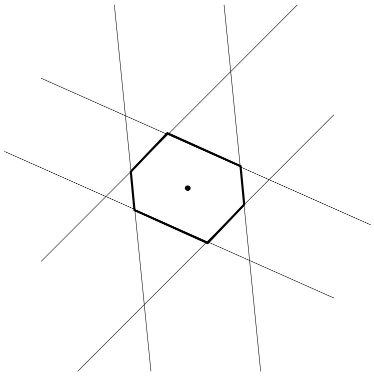

Figure 8. Any symmetric polytope is a section of a cube.

In Lecture 9 we shall see that all symmetric m-dimensional convex bodies have almost spherical sections of dimension about $\log m$. As one might expect, this is a great deal more difficult to prove for general bodies than just for the cube.

For the proof of Theorem 2.1, let's forget the symmetry assumption again and just ask for a polytope

$$K = \{x : \langle x, v_i \rangle \leq 1 \text{ for } 1 \leq i \leq m\}$$

with m facets for which

$$B_2^n \subset K \subset dB_2^n.$$

What do these inclusions say about the vectors (v_i)? The first implies that each v_i has length at most 1, since, if not, $v_i/|v_i|$ would be a vector in B_2^n but not in K. The second inclusion says that if x does not belong to dB_2^n then x does not belong to K: that is, if $|x| > d$, there is an i for which $\langle x, v_i \rangle > 1$. This is equivalent to the assertion that for every unit vector θ there is an i for which

$$\langle \theta, v_i \rangle \geq \frac{1}{d}\,.$$

Thus our problem is to look for as few vectors as possible, $v_1, v_2, \ldots, v_m$, of length at most 1, with the property that for every unit vector θ there is some v_i with $\langle \theta, v_i \rangle \geq 1/d$. It is clear that we cannot do better than to look for vectors of length exactly 1: that is, that we may assume that all facets of our polytope touch the ball. Henceforth we shall discuss only such vectors.

For a fixed unit vector v and $\varepsilon \in [0, 1)$, the set $C(\varepsilon, v)$ of $\theta \in S^{n-1}$ for which $\langle \theta, v \rangle \geq \varepsilon$ is called a *spherical cap* (or just a *cap*); when we want to be precise, we will call it the *ε-cap about v*. (Note that ε does not refer to the radius!) See Figure 9, left.

We want every $\theta \in S^{n-1}$ to belong to at least one of the $(1/d)$-caps determined by the (v_i). So our task is to estimate the number of caps of a given size needed to cover the entire sphere. The principal tool for doing this will be upper and lower estimates for the area of a spherical cap. As in the last lecture, we shall

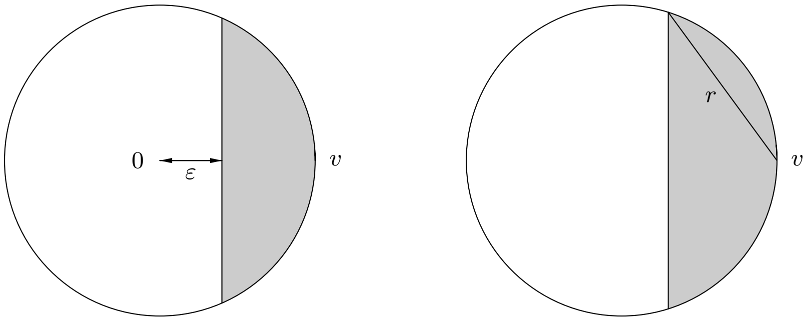

Figure 9. Left: ε-cap $C(\varepsilon, v)$ about v. Right: cap of radius r about v.

measure this area as a proportion of the sphere: that is, we shall use σ_{n-1} as our measure. Clearly, if we show that each cap occupies only a small proportion of the sphere, we can conclude that we need plenty of caps to cover the sphere. What is slightly more surprising is that once we have shown that spherical caps are not *too* small, we will also be able to deduce that we *can* cover the sphere without using too many.

In order to state the estimates for the areas of caps, it will sometimes be convenient to measure the size of a cap in terms of its radius, instead of using the ε measure. The cap of radius r about v is

$$\left\{\theta \in S^{n-1} : |\theta - v| \leq r\right\}$$

as illustrated in Figure 9, right. (In defining the radius of a cap in this way we are implicitly adopting a particular metric on the sphere: the metric induced by the usual Euclidean norm on $\mathbb{R}^n$.) The two estimates we shall use are given in the following lemmas.

LEMMA 2.2 (UPPER BOUND FOR SPHERICAL CAPS). *For $0 \leq \varepsilon < 1$, the cap $C(\varepsilon, u)$ on S^{n-1} has measure at most $e^{-n\varepsilon^2/2}$.*

LEMMA 2.3 (LOWER BOUND FOR SPHERICAL CAPS). *For $0 \leq r \leq 2$, a cap of radius r on S^{n-1} has measure at least $\frac{1}{2}(r/2)^{n-1}$.*

We can now prove Theorem 2.1.

PROOF. Lemma 2.2 implies the first assertion of Theorem 2.1 immediately. If K is a polytope in $\mathbb{R}^n$ with m facets and if $B_2^n \subset K \subset dB_2^n$, we can find m caps $C(\frac{1}{d}, v_i)$ covering S^{n-1}. Each covers at most $e^{-n/(2d^2)}$ of the sphere, so

$$m \geq \exp\left(\frac{n}{2d^2}\right).$$

To get the second assertion of Theorem 2.1 from Lemma 2.3 we need a little more argument. It suffices to find $m = 4^n$ points $v_1, v_2, \ldots, v_m$ on the sphere so that the caps of radius 1 centred at these points cover the sphere: see Figure 10. Such a set of points is called a 1-*net* for the sphere: every point of the sphere is

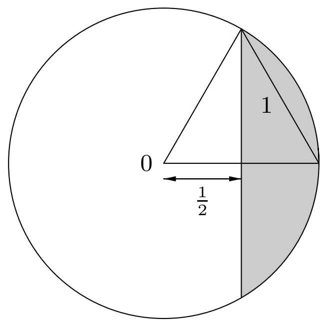

Figure 10. The $\frac{1}{2}$-cap has radius 1.

within distance 1 of some v_i. Now, if we choose a set of points on the sphere any two of which are at *least* distance 1 apart, this set cannot have too many points. (Such a set is called 1-*separated*.) The reason is that the caps of radius $\frac{1}{2}$ about these points will be disjoint, so the sum of their areas will be at most 1. A cap of radius $\frac{1}{2}$ has area at least $\left(\frac{1}{4}\right)^n$, so the number m of these caps satisfies $m \le 4^n$. This does the job, because a *maximal* 1-separated set is automatically a 1-net: if we can't add any further points that are at least distance 1 from all the points we have got, it can only be because every point of the sphere is *within* distance 1 of at least one of our chosen points. So the sphere has a 1-net consisting of only 4^n points, which is what we wanted to show. $\square$

Lemmas 2.2 and 2.3 are routine calculations that can be done in many ways. We leave Lemma 2.3 to the dedicated reader. Lemma 2.2, which will be quoted throughout Lectures 8 and 9, is closely related to the Gaussian decay of the volume of the ball described in the last lecture. At least for smallish ε (which is the interesting range) it can be proved as follows.

Proof. The proportion of S^{n-1} belonging to the cap $C(\varepsilon, u)$ equals the proportion of the solid ball that lies in the "spherical cone" illustrated in Figure 11.

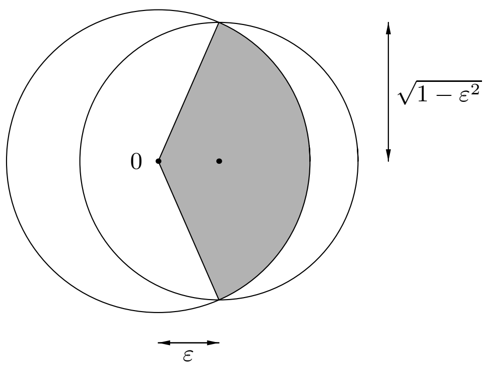

Figure 11. Estimating the area of a cap.

As is also illustrated, this spherical cone is contained in a ball of radius $\sqrt{1-\varepsilon^2}$ (if $\varepsilon \le 1/\sqrt{2}$), so the ratio of its volume to that of the ball is at most

$$\left(1-\varepsilon^2\right)^{n/2} \le e^{-n\varepsilon^2/2}. \qquad \square$$

In Lecture 8 we shall quote the upper estimate for areas of caps repeatedly. We shall in fact be using yet a third way to measure caps that differs very slightly from the $C(\varepsilon, u)$ description. The reader can easily check that the preceding argument yields the same estimate $e^{-n\varepsilon^2/2}$ for this other description.

Lecture 3. Fritz John's Theorem

In the first lecture we saw that the cube and the cross-polytope lie at distance at most $\sqrt{n}$ from the Euclidean ball in $\mathbb{R}^n$, and that for the simplex, the distance is at most n. It is intuitively clear that these estimates cannot be improved. In this lecture we shall describe a strong sense in which this is as bad as things get. The theorem we shall describe was proved by Fritz John [1948].

John considered ellipsoids inside convex bodies. If $(e_j)_1^n$ is an orthonormal basis of $\mathbb{R}^n$ and (α_j) are positive numbers, the ellipsoid

$$\left\{ x : \sum_1^n \frac{\langle x, e_j\rangle^2}{\alpha_j^2} \le 1 \right\}$$

has volume $v_n \prod \alpha_j$. John showed that each convex body contains a unique ellipsoid of largest volume and, more importantly, he *characterised* it. He showed that if K is a symmetric convex body in $\mathbb{R}^n$ and $\mathcal{E}$ is its maximal ellipsoid then

$$K \subset \sqrt{n}\,\mathcal{E}.$$

Hence, after an affine transformation (one taking $\mathcal{E}$ to B_2^n) we can arrange that

$$B_2^n \subset K \subset \sqrt{n} B_2^n.$$

A nonsymmetric K may require nB_2^n, like the simplex, rather than $\sqrt{n}B_2^n$.

John's characterisation of the maximal ellipsoid is best expressed after an affine transformation that takes the maximal ellipsoid to B_2^n. The theorem states that B_2^n is the maximal ellipsoid in K if a certain condition holds—roughly, that there be plenty of points of contact between the boundary of K and that of the ball. See Figure 12.

THEOREM 3.1 (JOHN'S THEOREM). *Each convex body K contains an unique ellipsoid of maximal volume. This ellipsoid is B_2^n if and only if the following conditions are satisfied: $B_2^n \subset K$ and (for some m) there are Euclidean unit vectors $(u_i)_1^m$ on the boundary of K and positive numbers $(c_i)_1^m$ satisfying*

$$\sum c_i u_i = 0 \tag{3.1}$$

and

$$\sum c_i \langle x, u_i\rangle^2 = |x|^2 \quad \text{for each } x \in \mathbb{R}^n. \tag{3.2}$$

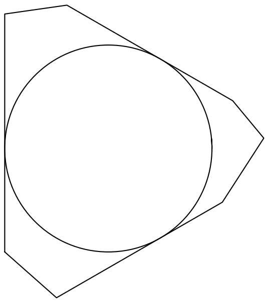

Figure 12. The maximal ellipsoid touches the boundary at many points.

According to the theorem the points at which the sphere touches ∂K can be given a mass distribution whose centre of mass is the origin and whose inertia tensor is the identity matrix. Let's see where these conditions come from. The first condition, (3.1), guarantees that the (u_i) do not all lie "on one side of the sphere". If they did, we could move the ball away from these contact points and blow it up a bit to obtain a larger ball in K. See Figure 13.

The second condition, (3.2), shows that the (u_i) behave rather like an orthonormal basis in that we can resolve the Euclidean norm as a (weighted) sum of squares of inner products. Condition (3.2) is equivalent to the statement that

$$x = \sum c_i \langle x, u_i \rangle u_i \quad \text{for all } x.$$

This guarantees that the (u_i) do not all lie close to a proper subspace of $\mathbb{R}^n$. If they did, we could shrink the ball a little in this subspace and expand it in an orthogonal direction, to obtain a larger ellipsoid inside K. See Figure 14.

Condition (3.2) is often written in matrix (or operator) notation as

$$\sum c_i \, u_i \otimes u_i = I_n \tag{3.3}$$

where I_n is the identity map on $\mathbb{R}^n$ and, for any unit vector u, $u \otimes u$ is the rank-one orthogonal projection onto the span of u, that is, the map $x \mapsto \langle x, u \rangle u$. The trace of such an orthogonal projection is 1. By equating the traces of the

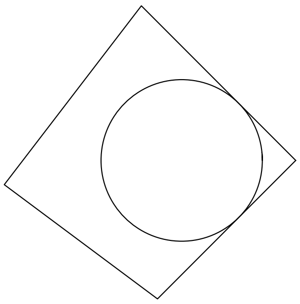

Figure 13. An ellipsoid where all contacts are on one side can grow.

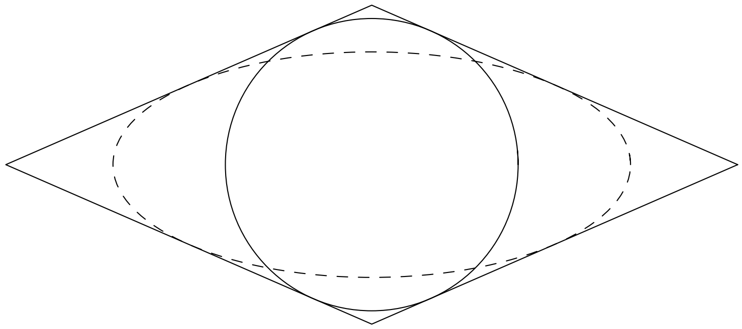

Figure 14. An ellipsoid (solid circle) whose contact points are all near one plane can grow.

matrices in the preceding equation, we obtain

$$\sum c_i = n.$$

In the case of a *symmetric* convex body, condition (3.1) is redundant, since we can take any sequence (u_i) of contact points satisfying condition (3.2) and replace each u_i by the pair $\pm u_i$ each with half the weight of the original.

Let's look at a few concrete examples. The simplest is the cube. For this body the maximal ellipsoid is B_2^n, as one would expect. The contact points are the standard basis vectors $(e_1, e_2, \ldots, e_n)$ of $\mathbb{R}^n$ and their negatives, and they satisfy

$$\sum_1^n e_i \otimes e_i = I_n.$$

That is, one can take all the weights c_i equal to 1 in (3.2). See Figure 15.

The simplest nonsymmetric example is the simplex. In general, there is no natural way to place a regular simplex in $\mathbb{R}^n$, so there is no natural description of the contact points. Usually the best way to talk about an n-dimensional simplex is to realise it in $\mathbb{R}^{n+1}$: for example as the convex hull of the standard basis

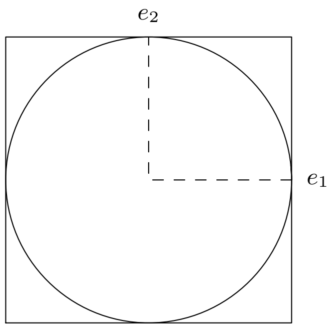

Figure 15. The maximal ellipsoid for the cube.

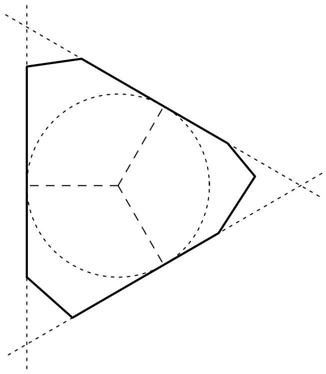

Figure 16. K is contained in the convex body determined by the hyperplanes tangent to the maximal ellipsoid at the contact points.

vectors in $\mathbb{R}^{n+1}$. We leave it as an exercise for the reader to come up with a nice description of the contact points.

One may get a bit more of a feel for the second condition in John's Theorem by interpreting it as a rigidity condition. A sequence of unit vectors (u_i) satisfying the condition (for some sequence (c_i)) has the property that if T is a linear map of determinant 1, not all the images Tu_i can have Euclidean norm less than 1.

John's characterisation immediately implies the inclusion mentioned earlier: if K is symmetric and $\mathcal{E}$ is its maximal ellipsoid then $K \subset \sqrt{n}\,\mathcal{E}$. To check this we may assume $\mathcal{E} = B_2^n$. At each contact point u_i, the convex bodies B_2^n and K have the same supporting hyperplane. For B_2^n, the supporting hyperplane at any point u is perpendicular to u. Thus if $x \in K$ we have $\langle x, u_i \rangle \le 1$ for each i, and we conclude that K is a subset of the convex body

$$C = \{x \in \mathbb{R}^n : \langle x, u_i \rangle \le 1 \text{ for } 1 \le i \le m\}. \tag{3.4}$$

An example of this is shown in Figure 16.

In the symmetric case, the same argument shows that for each $x \in K$ we have $|\langle x, u_i \rangle| \le 1$ for each i. Hence, for $x \in K$,

$$|x|^2 = \sum c_i \langle x, u_i \rangle^2 \le \sum c_i = n.$$

So $|x| \le \sqrt{n}$, which is exactly the statement $K \subset \sqrt{n}\,B_2^n$. We leave as a slightly trickier exercise the estimate $|x| \le n$ in the nonsymmetric case.

Proof of John's Theorem. The proof is in two parts, the harder of which is to show that if B_2^n is *an* ellipsoid of largest volume, then we can find an appropriate system of weights on the contact points. The easier part is to show that if such a system of weights exists, then B_2^n is the *unique* ellipsoid of maximal volume. We shall describe the proof only in the symmetric case, since the added complications in the general case add little to the ideas.

We begin with the easier part. Suppose there are unit vectors (u_i) in ∂K and numbers (c_i) satisfying

$$\sum c_i\, u_i \otimes u_i = I_n.$$

Let

$$\mathcal{E} = \left\{ x : \sum_1^n \frac{\langle x, e_j \rangle^2}{\alpha_j^2} \leq 1 \right\}$$

be an ellipsoid in K, for some orthonormal basis (e_j) and positive α_j. We want to show that

$$\prod_1^n \alpha_j \leq 1$$

and that the product is equal to 1 only if $\alpha_j = 1$ for all j.

Since $\mathcal{E} \subset K$ we have that for each i the hyperplane $\{x : \langle x, u_i \rangle = 1\}$ does not cut $\mathcal{E}$. This implies that each u_i belongs to the *polar ellipsoid*

$$\left\{ y : \sum_1^n \alpha_j^2 \langle y, e_j \rangle^2 \leq 1 \right\}.$$

(The reader unfamiliar with duality is invited to check this.) So, for each i,

$$\sum_{j=1}^n \alpha_j^2 \,\langle u_i, e_j \rangle^2 \leq 1.$$

Hence

$$\sum_i c_i \sum_j \alpha_j^2 \langle u_i, e_j \rangle^2 \leq \sum c_i = n.$$

But the left side of the equality is just $\sum_j \alpha_j^2$, because, by condition (3.2), we have

$$\sum_i c_i \,\langle u_i, e_j \rangle^2 = |e_j|^2 = 1$$

for each j. Finally, the fact that the geometric mean does not exceed the arithmetic mean (the AM/GM inequality) implies that

$$\left(\prod \alpha_j^2 \right)^{1/n} \leq \frac{1}{n} \sum \alpha_j^2 \leq 1,$$

and there is equality in the first of these inequalities only if all α_j are equal to 1.

We now move to the harder direction. Suppose B_2^n is an ellipsoid of largest volume in K. We want to show that there is a sequence of contact points (u_i) and positive weights (c_i) with

$$\frac{1}{n} I_n = \frac{1}{n} \sum c_i\, u_i \otimes u_i.$$

We already know that, if this is possible, we must have

$$\sum \frac{c_i}{n} = 1.$$

So our aim is to show that the matrix I_n/n can be written as a convex combination of (a finite number of) matrices of the form $u \otimes u$, where each u is a contact point. Since the space of matrices is finite-dimensional, the problem is simply to show that I_n/n belongs to the convex hull of the set of all such rank-one matrices,

$$T = \{u \otimes u : u \text{ is a contact point}\}.$$

We shall aim to get a contradiction by showing that if I_n/n is *not* in T, we can perturb the unit ball slightly to get a new ellipsoid in K of larger volume than the unit ball.

Suppose that I_n/n is not in T. Apply the separation theorem in the space of matrices to get a linear functional ϕ (on this space) with the property that

$$\phi\left(\frac{I_n}{n}\right) < \phi(u \otimes u) \tag{3.5}$$

for each contact point u. Observe that ϕ can be represented by an $n \times n$ matrix $H = (h_{jk})$, so that, for any matrix $A = (a_{jk})$,

$$\phi(A) = \sum_{jk} h_{jk} a_{jk}.$$

Since all the matrices $u \otimes u$ and I_n/n are symmetric, we may assume the same for H. Moreover, since these matrices all have the same trace, namely 1, the inequality $\phi(I_n/n) < \phi(u \otimes u)$ will remain unchanged if we add a constant to each diagonal entry of H. So we may assume that the trace of H is 0: but this says precisely that $\phi(I_n) = 0$.

Hence, unless the identity has the representation we want, we have found a symmetric matrix H with zero trace for which

$$\sum_{jk} h_{jk} (u \otimes u)_{jk} > 0$$

for every contact point u. We shall use this H to build a bigger ellipsoid inside K.

Now, for each vector u, the expression

$$\sum_{jk} h_{jk} (u \otimes u)_{jk}$$

is just the number $u^T H u$. For sufficiently small $\delta > 0$, the set

$$\mathcal{E}_\delta = \left\{x \in \mathbb{R}^n : x^T (I_n + \delta H) x \le 1\right\}$$

is an ellipsoid and as δ tends to 0 these ellipsoids approach B_2^n. If u is one of the original contact points, then

$$u^T (I_n + \delta H) u = 1 + \delta u^T H u > 1,$$

so u does not belong to $\mathcal{E}_\delta$. Since the boundary of K is compact (and the function $x \mapsto x^T H x$ is continuous) $\mathcal{E}_\delta$ will not contain any other point of ∂K as long as

δ is sufficiently small. Thus, for such δ, the ellipsoid $\mathcal{E}_\delta$ is strictly inside K and some slightly expanded ellipsoid is inside K.

It remains to check that each $\mathcal{E}_\delta$ has volume at least that of B_2^n. If we denote by (μ_j) the eigenvalues of the symmetric matrix $I_n + \delta H$, the volume of $\mathcal{E}_\delta$ is $v_n \big/ \prod \mu_j$, so the problem is to show that, for each δ, we have $\prod \mu_j \le 1$. What we know is that $\sum \mu_j$ is the trace of $I_n + \delta H$, which is n, since the trace of H is 0. So the AM/GM inequality again gives

$$\prod \mu_j^{1/n} \le \frac{1}{n} \sum \mu_j \le 1,$$

as required. □

There is an analogue of John's Theorem that characterises the unique ellipsoid of minimal volume containing a given convex body. (The characterisation is almost identical, guaranteeing a resolution of the identity in terms of contact points of the body and the Euclidean sphere.) This minimal volume ellipsoid theorem can be deduced directly from John's Theorem by duality. It follows that, for example, the ellipsoid of minimal volume containing the cube $[-1,1]^n$ is the obvious one: the ball of radius $\sqrt{n}$. It has been mentioned several times without proof that the distance of the cube from the Euclidean ball in $\mathbb{R}^n$ is exactly $\sqrt{n}$. We can now see this easily: the ellipsoid of minimal volume outside the cube has volume $\left(\sqrt{n}\right)^n$ times that of the ellipsoid of maximal volume inside the cube. So we cannot sandwich the cube between homothetic ellipsoids unless the outer one is at least $\sqrt{n}$ times the inner one.

We shall be using John's Theorem several times in the remaining lectures. At this point it is worth mentioning important extensions of the result. We can view John's Theorem as a description of those linear maps from Euclidean space to a normed space (whose unit ball is K) that have largest determinant, subject to the condition that they have norm at most 1: that is, that they map the Euclidean ball into K. There are many other norms that can be put on the space of linear maps. John's Theorem is the starting point for a general theory that builds ellipsoids related to convex bodies by maximising determinants subject to other constraints on linear maps. This theory played a crucial role in the development of convex geometry over the last 15 years. This development is described in detail in [Tomczak-Jaegermann 1988].

Lecture 4. Volume Ratios and Spherical Sections of the Octahedron

In the second lecture we saw that the n-dimensional cube has almost spherical sections of dimension about $\log n$ but not more. In this lecture we will examine the corresponding question for the n-dimensional cross-polytope B_1^n. In itself, this body is as far from the Euclidean ball as is the cube in $\mathbb{R}^n$: its distance from the ball, in the sense described in Lecture 2 is $\sqrt{n}$. Remarkably, however, it has almost spherical sections whose dimension is about $\frac{1}{2}n$. We shall deduce this from what is perhaps an even more surprising statement concerning intersections of bodies. Recall that B_1^n contains the Euclidean ball of radius $\frac{1}{\sqrt{n}}$. If U is an orthogonal transformation of $\mathbb{R}^n$ then UB_1^n also contains this ball and hence so does the intersection $B_1^n \cap UB_1^n$. But, whereas B_1^n does not lie in any Euclidean ball of radius less than 1, we have the following theorem [Kašin 1977]:

THEOREM 4.1. *For each n, there is an orthogonal transformation U for which the intersection $B_1^n \cap UB_1^n$ is contained in the Euclidean ball of radius $32/\sqrt{n}$ (and contains the ball of radius $1/\sqrt{n}$).*

(The constant 32 can easily be improved: the important point is that it is independent of the dimension n.) The theorem states that the intersection of just two copies of the n-dimensional octahedron may be approximately spherical. Notice that if we tried to approximate the Euclidean ball by intersecting rotated copies of the cube, we would need exponentially many in the dimension, because the cube has only $2n$ facets and our approximation needs exponentially many facets. In contrast, the octahedron has a much larger number of facets, 2^n: but of course we need to do a lot more than just count facets in order to prove Theorem 4.1. Before going any further we should perhaps remark that the cube has a property that corresponds to Theorem 4.1. If Q is the cube and U is the same orthogonal transformation as in the theorem, the convex hull

$$\operatorname{conv}(Q \cup UQ)$$

is at distance at most 32 from the Euclidean ball.

In spirit, the argument we shall use to establish Theorem 4.1 is Kašin's original one but, following [Szarek 1978], we isolate the main ingredient and we organise the proof along the lines of [Pisier 1989]. Some motivation may be helpful. The ellipsoid of maximal volume inside B_1^n is the Euclidean ball of radius $\frac{1}{\sqrt{n}}$. (See Figure 3.) There are 2^n points of contact between this ball and the boundary of B_1^n: namely, the points of the form

$$\left(\pm\frac{1}{n}, \pm\frac{1}{n}, \ldots, \pm\frac{1}{n}\right),$$

one in the middle of each facet of B_1^n. The vertices,

$$(\pm 1, 0, 0, \ldots, 0), \ldots, \; (0, 0, \ldots, 0, \pm 1),$$

are the points of B_1^n furthest from the origin. We are looking for a rotation UB_1^n whose facets chop off the spikes of B_1^n (or vice versa). So we want to know that the points of B_1^n at distance about $1/\sqrt{n}$ from the origin are fairly typical, while those at distance 1 are atypical.

For a unit vector $\theta \in S^{n-1}$, let $r(\theta)$ be the radius of B_1^n in the direction θ,

$$r(\theta) = \frac{1}{\|\theta\|_1} = \left(\sum_1^n |\theta_i|\right)^{-1}.$$

In the first lecture it was explained that the volume of B_1^n can be written

$$v_n \int_{S^{n-1}} r(\theta)^n \, d\sigma$$

and that it is equal to $2^n/n!$. Hence

$$\int_{S^{n-1}} r(\theta)^n \, d\sigma = \frac{2^n}{n! \, v_n} \le \left(\frac{2}{\sqrt{n}}\right)^n .$$

Since the average of $r(\theta)^n$ is at most $\left(2/\sqrt{n}\,\right)^n$, the value of $r(\theta)$ cannot often be much more than $2/\sqrt{n}$. This feature of B_1^n is captured in the following definition of Szarek.

Definition. Let K be a convex body in $\mathbb{R}^n$. The *volume ratio* of K is

$$\operatorname{vr}(K) = \left(\frac{\operatorname{vol}(K)}{\operatorname{vol}(\mathcal{E})}\right)^{1/n},$$

where $\mathcal{E}$ is the ellipsoid of maximal volume in K.

The preceding discussion shows that $\operatorname{vr}(B_1^n) \le 2$ for all n. Contrast this with the cube in $\mathbb{R}^n$, whose volume ratio is about $\sqrt{n}/2$. The only property of B_1^n that we shall use to prove Kašin's Theorem is that its volume ratio is at most 2. For convenience, we scale everything up by a factor of $\sqrt{n}$ and prove the following.

Theorem 4.2. *Let K be a symmetric convex body in $\mathbb{R}^n$ that contains the Euclidean unit ball B_2^n and for which*

$$\left(\frac{\operatorname{vol}(K)}{\operatorname{vol}(B_2^n)}\right)^{1/n} = R.$$

Then there is an orthogonal transformation U of $\mathbb{R}^n$ for which

$$K \cap UK \subset 8R^2 B_2^n .$$

Proof. It is convenient to work with the norm on $\mathbb{R}^n$ whose unit ball is K. Let $\|\cdot\|$ denote this norm and $|\cdot|$ the standard Euclidean norm. Notice that, since $B_2^n \subset K$, we have $\|x\| \le |x|$ for all $x \in \mathbb{R}^n$.

The radius of the body $K \cap UK$ in a given direction is the minimum of the radii of K and UK in that direction. So the norm corresponding to the body $K \cap UK$ is the *maximum* of the norms corresponding to K and UK. We need to find an orthogonal transformation U with the property that

$$\max\left(\|U\theta\|, \|\theta\|\right) \ge \frac{1}{8R^2}$$

for every $\theta \in S^{n-1}$. Since the maximum of two numbers is at least their average, it will suffice to find U with

$$\frac{\|U\theta\| + \|\theta\|}{2} \ge \frac{1}{8R^2} \quad \text{for all } \theta.$$

For each $x \in \mathbb{R}^n$ write $N(x)$ for the average $\frac{1}{2}(\|Ux\| + \|x\|)$. One sees immediately that N is a norm (that is, it satisfies the triangle inequality) and that $N(x) \le |x|$ for every x, since U preserves the Euclidean norm. We shall show in a moment that there is a U for which

$$\int_{S^{n-1}} \frac{1}{N(\theta)^{2n}} d\sigma \le R^{2n}. \tag{4.1}$$

This says that $N(\theta)$ is large on average: we want to deduce that it is large everywhere.

Let ϕ be a point of the sphere and write $N(\phi) = t$, for $0 < t \le 1$. The crucial point will be that, if θ is close to ϕ, then $N(\theta)$ cannot be much more than t. To be precise, if $|\theta - \phi| \le t$ then

$$N(\theta) \le N(\phi) + N(\theta - \phi) \le t + |\theta - \phi| \le 2t.$$

Hence, $N(\theta)$ is at most $2t$ for every θ in a spherical cap of radius t about ϕ. From Lemma 2.3 we know that this spherical cap has measure at least

$$\frac{1}{2}\left(\frac{t}{2}\right)^{n-1} \ge \left(\frac{t}{2}\right)^n.$$

So $1/N(\theta)^{2n}$ is at least $1/(2t)^{2n}$ on a set of measure at least $(t/2)^n$. Therefore

$$\int_{S^{n-1}} \frac{1}{N(\theta)^{2n}} d\sigma \ge \frac{1}{(2t)^{2n}} \left(\frac{t}{2}\right)^n = \frac{1}{2^{3n} t^n}.$$

By (4.1), the integral is at most R^{2n}, so $t \ge 1/(8R^2)$. Thus our arbitrary point ϕ satisfies

$$N(\phi) \ge \frac{1}{8R^2}.$$

It remains to find U satisfying (4.1). Now, for any θ, we have

$$N(\theta)^2 = \left(\frac{\|U\theta\| + \|\theta\|}{2}\right)^2 \ge \|U\theta\| \, \|\theta\|,$$

so it will suffice to find a U for which

$$\int_{S^{n-1}} \frac{1}{\|U\theta\|^n \, \|\theta\|^n} \, d\sigma \le R^{2n}. \tag{4.2}$$

The hypothesis on the volume of K can be written in terms of the norm as

$$\int_{S^{n-1}} \frac{1}{\|\theta\|^n} \, d\sigma = R^n.$$

The group of orthogonal transformations carries an invariant probability measure. This means that we can average a function over the group in a natural way. In particular, if f is a function on the sphere and θ is some point on the sphere, the average over orthogonal U of the value $f(U\theta)$ is just the average of f on the sphere: averaging over U mimics averaging over the sphere:

$$\text{ave}_U\, f(U\theta) = \int_{S^{n-1}} f(\phi)\, d\sigma(\phi).$$

Hence,

$$\begin{aligned}\text{ave}_U \int_{S^{n-1}} \frac{1}{\|U\theta\|^n.\|\theta\|^n}\, d\sigma(\theta) &= \int_{S^{n-1}} \left(\text{ave}_U \frac{1}{\|U\theta\|^n}\right) \frac{1}{\|\theta\|^n}\, d\sigma(\theta) \\ &= \int_{S^{n-1}} \left(\int_{S^{n-1}} \frac{1}{\|\phi\|^n} d\sigma(\phi)\right) \frac{1}{\|\theta\|^n}\, d\sigma(\theta) \\ &= \left(\int_{S^{n-1}} \frac{1}{\|\theta\|^n}\, d\sigma(\theta)\right)^2 = R^{2n}.\end{aligned}$$

Since the average over all U of the integral is at most R^{2n}, there is at least one U for which the integral is at most R^{2n}. This is exactly inequality (4.2). $\square$

The choice of U in the preceding proof is a random one. The proof does not in any way tell us how to find an explicit U for which the integral is small. In the case of a general body K, this is hardly surprising, since we are assuming nothing about how the volume of K is distributed. But, in view of the earlier remarks about facets of B_1^n chopping off spikes of UB_1^n, it is tempting to think that for the particular body B_1^n we might be able to write down an appropriate U explicitly. In two dimensions the best choice of U is obvious: we rotate the diamond through 45° and after intersection we have a regular octagon as shown in Figure 17.

The most natural way to try to generalise this to higher dimensions is to look for a U such that each vertex of UB_1^n is exactly aligned through the centre of a facet of B_1^n: that is, for each standard basis vector e_i of $\mathbb{R}^n$, Ue_i is a multiple of one of the vectors $(\pm\frac{1}{n}, \dots, \pm\frac{1}{n})$. (The multiple is $\sqrt{n}$ since Ue_i has length 1.) Thus we are looking for an $n \times n$ orthogonal matrix U each of whose entries is

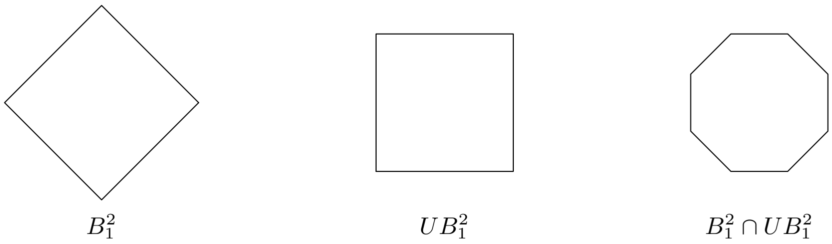

Figure 17. The best choice for U in two dimensions is a 45° rotation.

$\pm 1/\sqrt{n}$. Such matrices, apart from the factor $\sqrt{n}$, are called *Hadamard matrices*. In what dimensions do they exist? In dimensions 1 and 2 there are the obvious

$$(1) \quad \text{and} \quad \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

It is not too hard to show that in larger dimensions a Hadamard matrix cannot exist unless the dimension is a multiple of 4. It is an open problem to determine whether they exist in all dimensions that are multiples of 4. They are known to exist, for example, if the dimension is a power of 2: these examples are known as the Walsh matrices.

In spite of this, it seems extremely unlikely that one might prove Kašin's Theorem using Hadamard matrices. The Walsh matrices certainly do not give anything smaller than $n^{-1/4}$; pretty miserable compared with $n^{-1/2}$. There are some good reasons, related to Ramsey theory, for believing that one cannot expect to find genuinely explicit matrices of any kind that would give the right estimates.

Let's return to the question with which we opened the lecture and see how Theorem 4.1 yields almost spherical sections of octahedra. We shall show that, for each n, the $2n$-dimensional octahedron has an n-dimensional slice which is within distance 32 of the (n-dimensional) Euclidean ball. By applying the argument of the theorem to B_1^n, we obtain an $n \times n$ orthogonal matrix U such that

$$\|Ux\|_1 + \|x\|_1 \geq \frac{\sqrt{n}}{16}|x|$$

for every $x \in \mathbb{R}^n$, where $\|\cdot\|_1$ denotes the ℓ_1 norm. Now consider the map $T : \mathbb{R}^n \to \mathbb{R}^{2n}$ with matrix $\binom{U}{I}$. For each $x \in \mathbb{R}^n$, the norm of Tx in ℓ_1^{2n} is

$$\|Tx\|_1 = \|Ux\|_1 + \|x\|_1 \geq \frac{\sqrt{n}}{16}|x|.$$

On the other hand, the Euclidean norm of Tx is

$$|Tx| = \sqrt{|Ux|^2 + |x|^2} = \sqrt{2}\,|x|.$$

So, if y belongs to the image $T\mathbb{R}^n$, then, setting $y = Tx$,

$$\|y\|_1 \geq \frac{\sqrt{n}}{16}|x| = \frac{\sqrt{n}}{16\sqrt{2}}|y| = \frac{\sqrt{2n}}{32}|y|.$$

By the Cauchy–Schwarz inequality, we have $\|y\|_1 \leq \sqrt{2n}\,|y|$, so the slice of B_1^{2n} by the subspace $T\mathbb{R}^n$ has distance at most 32 from B_2^n, as we wished to show.

A good deal of work has been done on embedding of other subspaces of L_1 into ℓ_1-spaces of low dimension, and more generally subspaces of L_p into low-dimensional ℓ_p, for $1 < p < 2$. The techniques used come from probability theory: p-stable random variables, bootstrapping of probabilities and deviation estimates. We shall be looking at applications of the latter in Lectures 8 and 9.

The major references are [Johnson and Schechtman 1982; Bourgain et al. 1989; Talagrand 1990].

Volume ratios have been studied in considerable depth. They have been found to be closely related to the so-called cotype-2 property of normed spaces: this relationship is dealt with comprehensively in [Pisier 1989]. In particular, Bourgain and Milman [1987] showed that a bound for the cotype-2 constant of a space implies a bound for the volume ratio of its unit ball. This demonstrated, among other things, that there is a uniform bound for the volume ratios of slices of octahedra of all dimensions. A sharp version of this result was proved in [Ball 1991]: namely, that for each n, B_1^n has largest volume ratio among the balls of n-dimensional subspaces of L_1. The proof uses techniques that will be discussed in Lecture 6.

This is a good point to mention a result of Milman [1985] that looks superficially like the results of this lecture but lies a good deal deeper. We remarked that while we can almost get a sphere by intersecting two copies of B_1^n, this is very far from possible with two cubes. Conversely, we can get an almost spherical convex hull of two cubes but not of two copies of B_1^n. The *QS-Theorem* (an abbreviation for "quotient of a subspace") states that if we combine the two operations, intersection and convex hull, we can get a sphere no matter what body we start with.

THEOREM 4.3 (QS-THEOREM). *There is a constant M (independent of everything) such that, for any symmetric convex body K of any dimension, there are linear maps Q and S and an ellipsoid $\mathcal{E}$ with the following property: if $\tilde{K} = \operatorname{conv}(K \cup QK)$, then*

$$\mathcal{E} \subset \tilde{K} \cap S\tilde{K} \subset M\mathcal{E}.$$

Lecture 5. The Brunn–Minkowski Inequality and Its Extensions

In this lecture we shall introduce one of the most fundamental principles in convex geometry: the Brunn–Minkowski inequality. In order to motivate it, let's begin with a simple observation concerning convex sets in the plane. Let $K \subset \mathbb{R}^2$ be such a set and consider its slices by a family of parallel lines, for example those parallel to the y-axis. If the line $x = r$ meets K, call the length of the slice $v(r)$.

The graph of v is obtained by shaking K down onto the x-axis like a deck of cards (of different lengths). This is shown in Figure 18. It is easy to see that the function v is concave on its support. Towards the end of the last century, Brunn investigated what happens if we try something similar in higher dimensions.

Figure 19 shows an example in three dimensions. The central, hexagonal, slice has larger volume than the triangular slices at the ends: each triangular slice can be decomposed into four smaller triangles, while the hexagon is a union of six such triangles. So our first guess might be that the slice area is a concave

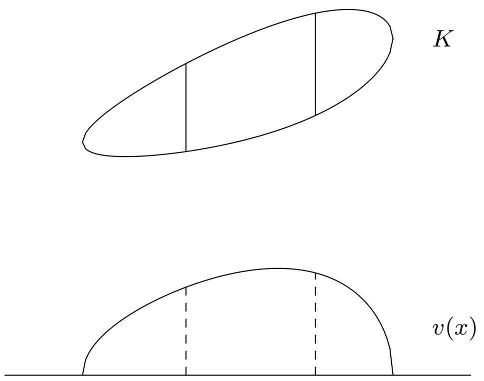

Figure 18. Shaking down a convex body.

function, just as slice length was concave for sets in the plane. That this is not always so can be seen by considering slices of a cone, parallel to its base: see Figure 20.

Since the area of a slice varies as the square of its distance from the cone's vertex, the area function obtained looks like a piece of the curve $y = x^2$, which is certainly not concave. However, it is reasonable to guess that the cone is an extremal example, since it is "only just" a convex body: its curved surface is "made up of straight lines". For this body, the square root of the slice function just manages to be concave on its support (since its graph is a line segment). So our second guess might be that for a convex body in $\mathbb{R}^3$, a slice-area function has a *square-root* that is concave on its support. This was proved by Brunn using an elegant symmetrisation method. His argument works just as well in higher dimensions to give the following result for the $(n-1)$-dimensional volumes of slices of a body in $\mathbb{R}^n$.

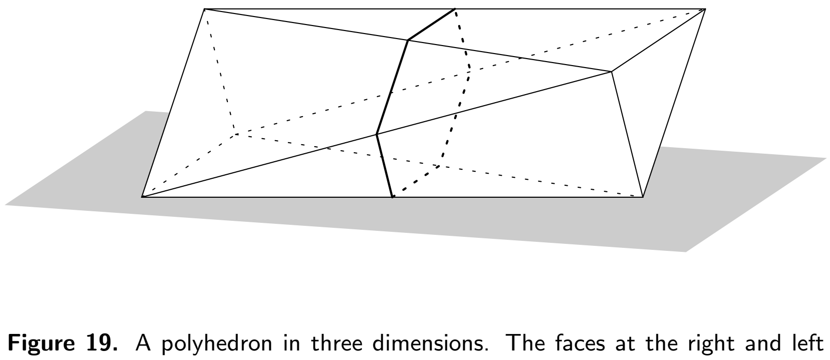

Figure 19. A polyhedron in three dimensions. The faces at the right and left are parallel.

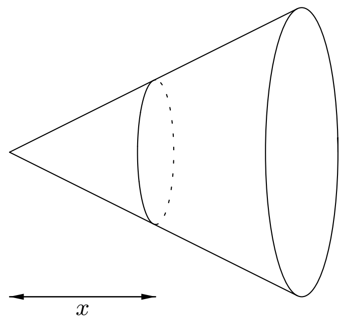

Figure 20. The area of a cone's section increases with x^2.

Theorem 5.1 (Brunn). *Let K be a convex body in $\mathbb{R}^n$, let u be a unit vector in $\mathbb{R}^n$, and for each r let H_r be the hyperplane*

$$\{x \in \mathbb{R}^n : \langle x, u \rangle = r\}.$$

Then the function

$$r \mapsto \mathrm{vol}\,(K \cap H_r)^{1/(n-1)}$$

is concave on its support.

One consequence of this is that if K is centrally symmetric, the largest slice perpendicular to a given direction is the central slice (since an even concave function is largest at 0). This is the situation in Figure 19.

Brunn's Theorem was turned from an elegant curiosity into a powerful tool by Minkowski. His reformulation works in the following way. Consider three parallel slices of a convex body in $\mathbb{R}^n$ at positions r, s and t, where $s = (1-\lambda)r + \lambda t$ for some $\lambda \in (0,1)$. This is shown in Figure 21.

Call the slices A_r, A_s, and A_t and think of them as subsets of $\mathbb{R}^{n-1}$. If $x \in A_r$ and $y \in A_t$, the point $(1-\lambda)x + \lambda y$ belongs to A_s: to see this, join the points (r,x) and (t,y) in $\mathbb{R}^n$ and observe that the resulting line segment crosses A_s at $(s,\,(1-\lambda)x + \lambda y)$. So A_s includes a new set

$$(1-\lambda)A_r + \lambda A_t := \{(1-\lambda)x + \lambda y : x \in A_r,\ y \in A_t\}.$$

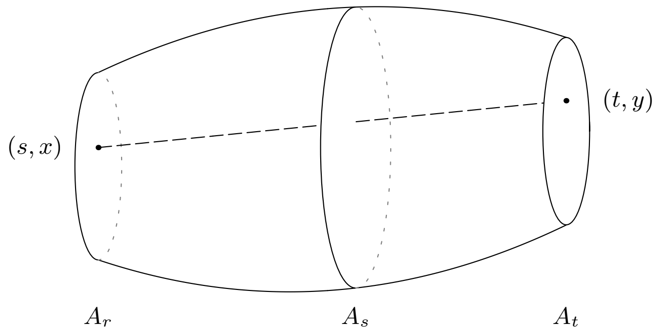

Figure 21. The section A_s contains the weighted average of A_r and A_t.

(This way of using the addition in $\mathbb{R}^n$ to define an addition of sets is called *Minkowski addition.*) Brunn's Theorem says that the volumes of the three sets A_r, A_s, and A_t in $\mathbb{R}^{n-1}$ satisfy

$$\operatorname{vol}(A_s)^{1/(n-1)} \geq (1-\lambda)\operatorname{vol}(A_r)^{1/(n-1)} + \lambda \operatorname{vol}(A_t)^{1/(n-1)}.$$

The Brunn–Minkowski inequality makes explicit the fact that all we really know about A_s is that it includes the Minkowski combination of A_r and A_t. Since we have now eliminated the role of the ambient space $\mathbb{R}^n$, it is natural to rewrite the inequality with n in place of $n-1$.

Theorem 5.2 (Brunn–Minkowski inequality). *If A and B are nonempty compact subsets of $\mathbb{R}^n$ then*

$$\operatorname{vol}((1-\lambda)A + \lambda B)^{1/n} \geq (1-\lambda)\operatorname{vol}(A)^{1/n} + \lambda \operatorname{vol}(B)^{1/n}.$$

(The hypothesis that A and B be nonempty corresponds in Brunn's Theorem to the restriction of a function to its support.) It should be remarked that the inequality is stated for general compact sets, whereas the early proofs gave the result only for convex sets. The first complete proof for the general case seems to be in [Lı̂usternik 1935].

To get a feel for the advantages of Minkowski's formulation, let's see how it implies the classical isoperimetric inequality in $\mathbb{R}^n$.

Theorem 5.3 (Isoperimetric inequality). *Among bodies of a given volume, Euclidean balls have least surface area.*

Proof. Let C be a compact set in $\mathbb{R}^n$ whose volume is equal to that of B_2^n, the Euclidean ball of radius 1. The surface "area" of C can be written

$$\operatorname{vol}(\partial C) = \lim_{\varepsilon \to 0} \frac{\operatorname{vol}(C + \varepsilon B_2^n) - \operatorname{vol}(C)}{\varepsilon},$$

as shown in Figure 22. By the Brunn–Minkowski inequality,

$$\operatorname{vol}(C + \varepsilon B_2^n)^{1/n} \geq \operatorname{vol}(C)^{1/n} + \varepsilon \operatorname{vol}(B_2^n)^{1/n}.$$

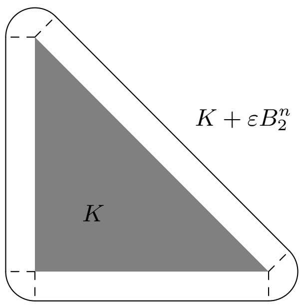

Figure 22. Expressing the area as a limit of volume increments.

Hence

$$\begin{aligned}\operatorname{vol}(C+\varepsilon B_2^n) &\geq \left(\operatorname{vol}(C)^{1/n}+\varepsilon \operatorname{vol}(B_2^n)^{1/n}\right)^n \\ &\geq \operatorname{vol}(C)+n\varepsilon \operatorname{vol}(C)^{(n-1)/n}\operatorname{vol}(B_2^n)^{1/n}.\end{aligned}$$

So

$$\operatorname{vol}(\partial C) \geq n \operatorname{vol}(C)^{(n-1)/n}\operatorname{vol}(B_2^n)^{1/n}.$$

Since C and B_2^n have the same volume, this shows that $\operatorname{vol}(\partial C) \geq n \operatorname{vol}(B_2^n)$, and the latter equals $\operatorname{vol}(\partial B_2^n)$, as we saw in Lecture 1. □

This relationship between the Brunn–Minkowski inequality and the isoperimetric inequality will be explored in a more general context in Lecture 8.

The Brunn–Minkowski inequality has an alternative version that is formally weaker. The AM/GM inequality shows that, for λ in $(0,1)$,

$$(1-\lambda)\operatorname{vol}(A)^{1/n}+\lambda \operatorname{vol}(B)^{1/n} \geq \operatorname{vol}(A)^{(1-\lambda)/n}\operatorname{vol}(B)^{\lambda/n}.$$

So the Brunn–Minkowski inequality implies that, for compact sets A and B and $\lambda \in (0,1)$,

$$\operatorname{vol}((1-\lambda)A+\lambda B) \geq \operatorname{vol}(A)^{1-\lambda}\operatorname{vol}(B)^{\lambda}. \tag{5.1}$$

Although this multiplicative Brunn–Minkowski inequality is weaker than the Brunn–Minkowski inequality for *particular* A, B, and λ, if one knows (5.1) for *all* A, B, and λ one can easily deduce the Brunn–Minkowski inequality for all A, B, and λ. This deduction will be left for the reader.

Inequality (5.1) has certain advantages over the Brunn–Minkowski inequality.

(i) We no longer need to stipulate that A and B be nonempty, which makes the inequality easier to use.
(ii) The dimension n has disappeared.
(iii) As we shall see, the multiplicative inequality lends itself to a particularly simple proof because it has a generalisation from sets to functions.

Before we describe the functional Brunn–Minkowski inequality let's just remark that the multiplicative Brunn–Minkowski inequality can be reinterpreted back in the setting of Brunn's Theorem: if $r \mapsto v(r)$ is a function obtained by scanning a convex body with parallel hyperplanes, then $\log v$ is a concave function (with the usual convention regarding $-\infty$).

In order to move toward a functional generalisation of the multiplicative Brunn–Minkowski inequality let's reinterpret inequality (5.1) in terms of the characteristic functions of the sets involved. Let f, g, and m denote the characteristic functions of A, B, and $(1-\lambda)A+\lambda B$ respectively; so, for example, $f(x)=1$ if $x \in A$ and 0 otherwise. The volumes of A, B, and $(1-\lambda)A+\lambda B$ are the integrals $\int_{\mathbb{R}^n} f$, $\int_{\mathbb{R}^n} g$, and $\int_{\mathbb{R}^n} m$. The Brunn–Minkowski inequality says that

$$\int m \geq \left(\int f\right)^{1-\lambda}\left(\int g\right)^{\lambda}.$$

But what is the relationship between f, g, and m that guarantees its truth? If $f(x) = 1$ and $g(y) = 1$ then $x \in A$ and $y \in B$, so

$$(1-\lambda)x + \lambda y \in (1-\lambda)A + \lambda B,$$

and hence $m\left((1-\lambda)x + \lambda y\right) = 1$. This certainly ensures that

$$m\left((1-\lambda)x + \lambda y\right) \geq f(x)^{1-\lambda} g(y)^{\lambda} \quad \text{for any } x \text{ and } y \text{ in } \mathbb{R}^n.$$

This inequality for the three functions at least has a homogeneity that matches the desired inequality for the integrals. In a series of papers, Prékopa and Leindler proved that this homogeneity is enough.

THEOREM 5.4 (THE PRÉKOPA–LEINDLER INEQUALITY). *If f, g and m are nonnegative measurable functions on $\mathbb{R}^n$, $\lambda \in (0,1)$ and for all x and y in $\mathbb{R}^n$,*

$$m\left((1-\lambda)x + \lambda y\right) \geq f(x)^{1-\lambda} g(y)^{\lambda} \tag{5.2}$$

then

$$\int m \geq \left(\int f\right)^{1-\lambda} \left(\int g\right)^{\lambda}.$$

It is perhaps helpful to notice that the Prékopa–Leindler inequality looks like Hölder's inequality, backwards. If f and g were given and we set

$$m(z) = f(z)^{1-\lambda} g(z)^{\lambda}$$

(for each z), then Hölder's inequality says that

$$\int m \leq \left(\int f\right)^{1-\lambda} \left(\int g\right)^{\lambda}.$$

(Hölder's inequality is often written with $1/p$ instead of $1-\lambda$, $1/q$ instead of λ, and f, g replaced by F^p, G^q.) The difference between Prékopa–Leindler and Hölder is that, in the former, the value $m(z)$ may be much larger since it is a supremum over many pairs (x, y) satisfying $z = (1-\lambda)x + \lambda y$ rather than just the pair (z, z).

Though it generalises the Brunn–Minkowski inequality, the Prékopa–Leindler inequality is a good deal simpler to prove, once properly formulated. The argument we shall use seems to have appeared first in [Brascamp and Lieb 1976b]. The crucial point is that the passage from sets to functions allows us to prove the inequality by induction on the dimension, using only the one-dimensional case. We pay the small price of having to do a bit extra for this case.

PROOF OF THE PRÉKOPA–LEINDLER INEQUALITY. We start by checking the one-dimensional Brunn–Minkowski inequality. Suppose A and B are nonempty measurable subsets of the line. Using $|\cdot|$ to denote length, we want to show that

$$|(1-\lambda)A + \lambda B| \geq (1-\lambda)|A| + \lambda |B|.$$

We may assume that A and B are compact and we may shift them so that the right-hand end of A and the left-hand end of B are both at 0. The set $(1-\lambda)A + \lambda B$ now includes the essentially disjoint sets $(1-\lambda)A$ and λB, so its length is at least the sum of the lengths of these sets.

Now suppose we have nonnegative integrable functions f, g, and m on the line, satisfying condition (5.2). We may assume that f and g are bounded. Since the inequality to be proved has the same homogeneity as the hypothesis (5.2), we may also assume that f and g are normalised so that $\sup f = \sup g = 1$. By Fubini's Theorem, we can write the integrals of f and g as integrals of the lengths of their level sets:

$$\int f(x)\,dx = \int_0^1 |(f \geq t)|\;dt,$$

and similarly for g. If $f(x) \geq t$ and $g(y) \geq t$ then $m\left((1-\lambda)x + \lambda y\right) \geq t$. So we have the inclusion

$$(m \geq t) \supset (1-\lambda)(f \geq t) + \lambda(g \geq t).$$

For $0 \leq t < 1$ the sets on the right are nonempty so the one-dimensional Brunn–Minkowski inequality shows that

$$|(m \geq t)| \geq (1-\lambda)\,|(f \geq t)| + \lambda\,|(g \geq t)|\,.$$

Integrating this inequality from 0 to 1 we get

$$\int m \geq (1-\lambda)\int f + \lambda \int g,$$

and the latter expression is at least

$$\left(\int f\right)^{1-\lambda}\left(\int g\right)^{\lambda}$$

by the AM/GM inequality. This does the one-dimensional case.

The induction that takes us into higher dimensions is quite straightforward, so we shall just sketch the argument for sets in $\mathbb{R}^n$, rather than functions. Suppose A and B are two such sets and, for convenience, write

$$C = (1-\lambda)A + \lambda B.$$

Choose a unit vector u and, as before, let H_r be the hyperplane

$$\{x \in \mathbb{R}^n : \langle x, u\rangle = r\}$$

perpendicular to u at "position" r. Let A_r denote the slice $A \cap H_r$ and similarly for B and C, and regard these as subsets of $\mathbb{R}^{n-1}$. If r and t are real numbers, and if $s = (1-\lambda)r + \lambda t$, the slice C_s includes $(1-\lambda)A_r + \lambda B_t$. (This is reminiscent of the earlier argument relating Brunn's Theorem to Minkowski's reformulation.) By the inductive hypothesis in $\mathbb{R}^{n-1}$,

$$\operatorname{vol}(C_s) \geq \operatorname{vol}(A_r)^{1-\lambda}.\operatorname{vol}(B_t)^{\lambda}.$$

Let f, g, and m be the functions on the line, given by

$$f(x) = \operatorname{vol}(A_x), \quad g(x) = \operatorname{vol}(B_x), \quad m(x) = \operatorname{vol}(C_x).$$

Then, for r, s, and t as above,

$$m(s) \geq f(r)^{1-\lambda} g(t)^{\lambda}.$$

By the one-dimensional Prékopa–Leindler inequality,

$$\int m \geq \left(\int f\right)^{1-\lambda} \left(\int g\right)^{\lambda}.$$

But this is exactly the statement $\operatorname{vol}(C) \geq \operatorname{vol}(A)^{1-\lambda} \operatorname{vol}(B)^{\lambda}$, so the inductive step is complete. □

The proof illustrates clearly why the Prékopa–Leindler inequality makes things go smoothly. Although we only carried out the induction for *sets*, we required the one-dimensional result for the *functions* we get by scanning sets in $\mathbb{R}^n$.

To close this lecture we remark that the Brunn–Minkowski inequality has numerous extensions and variations, not only in convex geometry, but in combinatorics and information theory as well. One of the most surprising and delightful is a theorem of Busemann [1949].

Theorem 5.5 (Busemann). *Let K be a symmetric convex body in $\mathbb{R}^n$, and for each unit vector u let $r(u)$ be the volume of the slice of K by the subspace orthogonal to u. Then the body whose radius in each direction u is $r(u)$ is itself convex.*

The Brunn–Minkowski inequality is the starting point for a highly developed classical theory of convex geometry. We shall barely touch upon the theory in these notes. A comprehensive reference is the recent book [Schneider 1993].

Lecture 6. Convolutions and Volume Ratios: The Reverse Isoperimetric Problem

In the last lecture we saw how to deduce the classical isoperimetric inequality in $\mathbb{R}^n$ from the Brunn–Minkowski inequality. In this lecture we will answer the reverse question. This has to be phrased a bit carefully, since there is no upper limit to the surface area of a body of given volume, even if we restrict attention to convex bodies. (Consider a very thin pancake.) For this reason it is natural to consider affine equivalence classes of convex bodies, and the question becomes: given a convex body, how small can we make its surface area by applying an affine (or linear) transformation that preserves volume? The answer is provided by the following theorem from [Ball 1991].

THEOREM 6.1. *Let K be a convex body and T a regular solid simplex in $\mathbb{R}^n$. Then there is an affine image of K whose volume is the same as that of T and whose surface area is no larger than that of T.*

Thus, modulo affine transformations, simplices have the largest surface area among convex bodies of a given volume. If K is assumed to be centrally symmetric then the estimate can be strengthened: the cube is extremal among symmetric bodies. A detailed proof of Theorem 6.1 would be too long for these notes. We shall instead describe how the symmetric case is proved, since this is considerably easier but illustrates the most important ideas.

Theorem 6.1 and the symmetric analogue are both deduced from volume-ratio estimates. In the latter case the statement is that among symmetric convex bodies, the cube has largest volume ratio. Let's see why this solves the reverse isoperimetric problem. If Q is any cube, the surface area and volume of Q are related by

$$\mathrm{vol}(\partial Q) = 2n\,\mathrm{vol}(Q)^{(n-1)/n}.$$

We wish to show that any other convex body K has an affine image $\tilde{K}$ for which

$$\mathrm{vol}(\partial \tilde{K}) \le 2n\,\mathrm{vol}(\tilde{K})^{(n-1)/n}.$$

Choose $\tilde{K}$ so that its maximal volume ellipsoid is B_2^n, the Euclidean ball of radius 1. The volume of $\tilde{K}$ is then at most 2^n, since this is the volume of the cube whose maximal ellipsoid is B_2^n. As in the previous lecture,

$$\mathrm{vol}(\partial \tilde{K}) = \lim_{\varepsilon\to 0} \frac{\mathrm{vol}(\tilde{K}+\varepsilon B_2^n) - \mathrm{vol}(\tilde{K})}{\varepsilon}.$$

Since $\tilde{K} \supset B_2^n$, the second expression is at most

$$\begin{aligned}\lim_{\varepsilon\to 0} \frac{\mathrm{vol}(\tilde{K}+\varepsilon \tilde{K}) - \mathrm{vol}(\tilde{K})}{\varepsilon} &= \mathrm{vol}(\tilde{K}) \lim_{\varepsilon\to 0} \frac{(1+\varepsilon)^n - 1}{\varepsilon} \\ &= n\,\mathrm{vol}(\tilde{K}) = n\,\mathrm{vol}(\tilde{K})^{1/n}\,\mathrm{vol}(\tilde{K})^{(n-1)/n} \\ &\le 2n\,\mathrm{vol}(\tilde{K})^{(n-1)/n},\end{aligned}$$

which is exactly what we wanted.

The rest of this lecture will thus be devoted to explaining the proof of the volume-ratio estimate:

THEOREM 6.2. *Among symmetric convex bodies the cube has largest volume ratio.*

As one might expect, the proof of Theorem 6.2 makes use of John's Theorem from Lecture 3. The problem is to show that, if K is a convex body whose maximal ellipsoid is B_2^n, then $\mathrm{vol}(K) \le 2^n$. As we saw, it is a consequence of

John's theorem that if B_2^n is the maximal ellipsoid in K, there is a sequence (u_i) of unit vectors and a sequence (c_i) of positive numbers for which

$$\sum c_i\, u_i \otimes u_i = I_n$$

and for which

$$K \subset C := \{x : |\langle x, u_i\rangle| \le 1 \text{ for } 1 \le i \le m\}\,.$$

We shall show that this C has volume at most 2^n. The principal tool will be a sharp inequality for norms of generalised convolutions. Before stating this let's explain some standard terms from harmonic analysis.

If f and $g : \mathbb{R} \to \mathbb{R}$ are bounded, integrable functions, we define the *convolution* $f * g$ of f and g by

$$f * g(x) = \int_{\mathbb{R}} f(y) g(x - y)\, dy.$$

Convolutions crop up in many areas of mathematics and physics, and a good deal is known about how they behave. One of the most fundamental inequalities for convolutions is Young's inequality: If $f \in L_p$, $g \in L_q$, and

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{s},$$

then

$$\|f * g\|_s \le \|f\|_p \, \|g\|_q\,.$$

(Here $\|\cdot\|_p$ means the L_p norm on $\mathbb{R}$, and so on.) Once we have Young's inequality, we can give a meaning to convolutions of functions that are not both integrable and bounded, provided that they lie in the correct L_p spaces. Young's inequality holds for convolution on any locally compact group, for example the circle. On compact groups it is sharp: there is equality for constant functions. But on $\mathbb{R}$, where constant functions are not integrable, the inequality can be improved (for most values of p and q). It was shown by Beckner [1975] and Brascamp and Lieb [1976a] that the correct constant in Young's inequality is attained if f and g are appropriate Gaussian densities: that is, for some positive a and b, $f(t) = e^{-at^2}$ and $g(t) = e^{-bt^2}$. (The appropriate choices of a and b and the value of the best constant for each p and q will not be stated here. Later we shall see that they can be avoided.)

How are convolutions related to convex bodies? To answer this question we need to rewrite Young's inequality slightly. If $1/r + 1/s = 1$, the L_s norm $\|f * g\|_s$ can be realised as

$$\int_{\mathbb{R}} (f * g)(x) h(x)$$

for some function h with $\|h\|_r = 1$. So the inequality says that, if $1/p + 1/q + 1/r = 2$, then

$$\iint f(y) g(x - y) h(x)\, dy\, dx \le \|f\|_p \, \|g\|_q \, \|h\|_r\,.$$

We may rewrite the inequality again with $h(-x)$ in place of $h(x)$, since this doesn't affect $\|h\|_r$:

$$\iint f(y)g(x-y)h(-x)\,dy\,dx \le \|f\|_p\,\|g\|_q\,\|h\|_r\,. \tag{6.1}$$

This can be written in a more symmetric form via the map from $\mathbb{R}^2$ into $\mathbb{R}^3$ that takes (x,y) to $(y, x-y, -x) =: (u,v,w)$. The range of this map is the subspace

$$H = \{(u,v,w) : u+v+w = 0\}\,.$$

Apart from a factor coming from the Jacobian of this map, the integral can be written

$$\int_H f(u)g(v)h(w),$$

where the integral is with respect to two-dimensional measure on the subspace H. So Young's inequality and its sharp forms estimate the integral of a product function on $\mathbb{R}^3$ over a subspace. What is the simplest product function? If f, g, and h are each the characteristic function of the interval $[-1,1]$, the function F given by

$$F(u,v,w) = f(u)g(v)h(w)$$

is the characteristic function of the cube $[-1,1]^3 \subset \mathbb{R}^3$. The integral of F over a subspace of $\mathbb{R}^3$ is thus the area of a slice of the cube: the area of a certain convex body. So there is some hope that we might use a convolution inequality to estimate volumes.

Brascamp and Lieb proved rather more than the sharp form of Young's inequality stated earlier. They considered not just two-dimensional subspaces of $\mathbb{R}^3$ but n-dimensional subspaces of $\mathbb{R}^m$. It will be more convenient to state their result using expressions analogous to those in (6.1) rather than using integrals over subspaces. Notice that the integral

$$\iint f(y)g(x-y)h(-x)\,dy\,dx$$

can be written

$$\int_{\mathbb{R}^2} f\big(\langle x, v_1\rangle\big)g\big(\langle x, v_2\rangle\big)h\big(\langle x, v_3\rangle\big)\,dx,$$

where $v_1 = (0,1)$, $v_2 = (1,-1)$ and $v_3 = (-1,0)$ are vectors in $\mathbb{R}^2$. The theorem of Brascamp and Lieb is the following.

Theorem 6.3. *If $(v_i)_1^m$ are vectors in $\mathbb{R}^n$ and $(p_i)_1^m$ are positive numbers satisfying*

$$\sum_1^m \frac{1}{p_i} = n,$$

and if $(f_i)_1^m$ are nonnegative measurable functions on the line, then

$$\frac{\int_{\mathbb{R}^n}\prod_1^m f_i\left(\langle x, v_i\rangle\right)}{\prod_1^m \|f_i\|_{p_i}}$$

is "maximised" when the (f_i) *are appropriate Gaussian densities*: $f_i(t) = e^{-a_i t^2}$, *where the* a_i *depend upon* m, n, *the* p_i, *and the* v_i.

The word maximised is in quotation marks since there are degenerate cases for which the maximum is not attained. The value of the maximum is not easily computed since the a_i are the solutions of nonlinear equations in the p_i and v_i. This apparently unpleasant problem evaporates in the context of convex geometry: the inequality has a normalised form, introduced in [Ball 1990], which fits perfectly with John's Theorem.

THEOREM 6.4. *If* $(u_i)_1^m$ *are unit vectors in* $\mathbb{R}^n$ *and* $(c_i)_1^m$ *are positive numbers for which*

$$\sum_1^m c_i\, u_i \otimes u_i = I_n,$$

and if $(f_i)_1^m$ *are nonnegative measurable functions*, *then*

$$\int_{\mathbb{R}^n} \prod f_i\left(\langle x, u_i\rangle\right)^{c_i} \le \prod \left(\int f_i\right)^{c_i}.$$

In this reformulation of the theorem, the c_i play the role of $1/p_i$: the Fritz John condition ensures that $\sum c_i = n$ as required, and miraculously guarantees that the correct constant in the inequality is 1 (as written). The functions f_i have been replaced by $f_i^{c_i}$, since this ensures that equality occurs if the f_i are identical Gaussian densities. It may be helpful to see why this is so. If $f_i(t) = e^{-t^2}$ for all i, then

$$\prod f_i\left(\langle x, u_i\rangle\right)^{c_i} = \exp\Bigl(-\sum c_i \langle x, u_i\rangle^2\Bigr) = e^{-|x|^2} = \prod_1^n e^{-x_i^2},$$

so the integral is

$$\left(\int e^{-t^2}\right)^n = \prod\left(\int e^{-t^2}\right)^{c_i} = \prod\left(\int f_i\right)^{c_i}.$$

Armed with Theorem 6.4, let's now prove Theorem 6.2.

PROOF OF THE VOLUME-RATIO ESTIMATE. Recall that our aim is to show that, for u_i and c_i as usual, the body

$$C = \{x : |\langle x, u_i\rangle| \le 1 \text{ for } 1 \le i \le m\}$$

has volume at most 2^n. For each i let f_i be the characteristic function of the interval $[-1, 1]$ in $\mathbb{R}$. Then the function

$$x \mapsto \prod f_i\bigl(\langle x, u_i\rangle\bigr)^{c_i}$$

is exactly the characteristic function of C. Integrating and applying Theorem 6.4 we have

$$\operatorname{vol}(C) \leq \prod \left(\int f_i \right)^{c_i} = \prod 2^{c_i} = 2^n. \qquad \square$$

The theorems of Brascamp and Lieb and Beckner have been greatly extended over the last twenty years. The main result in Beckner's paper solved the old problem of determining the norm of the Fourier transform between L_p spaces. There are many classical inequalities in harmonic analysis for which the best constants are now known. The paper [Lieb 1990] contains some of the most up-to-date discoveries and gives a survey of the history of these developments.

The methods described here have many other applications to convex geometry. There is also a reverse form of the Brascamp–Lieb inequality appropriate for analysing, for example, the ratio of the volume of a body to that of the minimal ellipsoid containing it.

Lecture 7. The Central Limit Theorem and Large Deviation Inequalities

The material in this short lecture is not really convex geometry, but is intended to provide a context for what follows. For the sake of readers who may not be familiar with probability theory, we also include a few words about independent random variables.

To begin with, a *probability measure* μ on a set Ω is just a measure of total mass $\mu(\Omega) = 1$. Real-valued functions on Ω are called random variables and the integral of such a function $X : \Omega \to \mathbb{R}$, its mean, is written EX and called the expectation of X. The variance of X is $E(X - EX)^2$. It is customary to suppress the reference to Ω when writing the measures of sets defined by random variables. Thus

$$\mu(\{\omega \in \Omega : X(\omega) < 1\})$$

is written $\mu(X < 1)$: the probability that X is less than 1.

Two crucial, and closely related, ideas distinguish probability theory from general measure theory. The first is independence. Two random variables X and Y are said to be independent if, for any functions f and g,

$$Ef(X)g(Y) = Ef(X)\,Eg(Y).$$

Independence can always be viewed in a canonical way. Let (Ω, μ) be a product space $(\Omega_1 \times \Omega_2,\ \mu_1 \otimes \mu_2)$, where μ_1 and μ_2 are probabilities. Suppose X and Y are random variables on Ω for which the value $X(\omega_1, \omega_2)$ depends only upon ω_1 while $Y(\omega_1, \omega_2)$ depends only upon ω_2. Then any integral (that converges appropriately)

$$Ef(X)g(Y) = \int f(X(s))\,g(Y(t))\,d\mu_1 \otimes \mu_2(s,t)$$

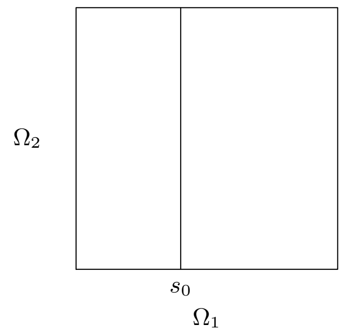

Figure 23. Independence and product spaces

can be written as the product of integrals

$$\int f(X(s))\,d\mu_1(s)\int g(Y(t))\,d\mu_2(t)=Ef(X)Eg(Y)$$

by Fubini's Theorem. Putting it another way, on each line $\{(s_0,t):t\in\Omega_2\}$, X is fixed, while Y exhibits its full range of behaviour in the correct proportions. This is illustrated in Figure 23.

In a similar way, a sequence $X_1, X_2, \ldots, X_n$ of independent random variables arises if each variable is defined on the product space $\Omega_1\times\Omega_2\times\ldots\times\Omega_n$ and X_i depends only upon the i-th coordinate.

The second crucial idea, which we will not discuss in any depth, is the use of many different σ-fields on the same space. The simplest example has already been touched upon. The product space $\Omega_1\times\Omega_2$ carries two σ-fields, much smaller than the product field, which it inherits from Ω_1 and Ω_2 respectively. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are the σ-fields on Ω_1 and Ω_2, the sets of the form $A\times\Omega_2\subset\Omega_1\times\Omega_2$ for $A\in\mathcal{F}_1$ form a σ-field on $\Omega_1\times\Omega_2$; let's call it $\tilde{\mathcal{F}}_1$. Similarly,

$$\tilde{\mathcal{F}}_2=\{\Omega_1\times B : B\in\mathcal{F}_2\}.$$

"Typical" members of these σ-fields are shown in Figure 24.

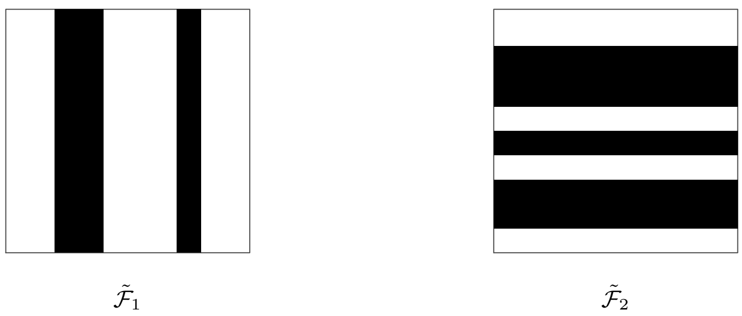

Figure 24. Members of the "small" σ-fields $\tilde{\mathcal{F}}_1$ and $\tilde{\mathcal{F}}_2$ on $\Omega_1\times\Omega_2$.

One of the most beautiful and significant principles in mathematics is the central limit theorem: any random quantity that arises as the sum of many small independent contributions is distributed very much like a Gaussian random variable. The most familiar example is coin tossing. We use a coin whose decoration is a bit austere: it has $+1$ on one side and -1 on the other. Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ be the outcomes of n independent tosses. Thus the ε_i are independent random variables, each of which takes the values $+1$ and -1 each with probability $\frac{1}{2}$. (Such random variables are said to have a *Bernoulli distribution.*) Then the normalised sum

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

belongs to an interval I of the real line with probability very close to

$$\frac{1}{\sqrt{2\pi}} \int_I e^{-t^2/2}\, dt.$$

The normalisation $1/\sqrt{n}$, ensures that the variance of S_n is 1: so there is some hope that the S_n will all be similarly distributed.

The standard proof of the central limit theorem shows that much more is true. Any sum of the form

$$\sum_1^n a_i \varepsilon_i$$

with real coefficients a_i will have a roughly Gaussian distribution as long as each a_i is fairly small compared with $\sum {a_i}^2$. Some such smallness condition is clearly needed since if

$$a_1 = 1 \quad \text{and} \quad a_2 = a_3 = \cdots = a_n = 0,$$

the sum is just ε_1, which is not much like a Gaussian. However, in many instances, what one really wants is not that the sum is distributed like a Gaussian, but merely that the sum cannot be large (or far from average) much more often than an appropriate Gaussian variable. The example above clearly satisfies such a condition: ε_1 never deviates from its mean, 0, by more than 1.

The following inequality provides a deviation estimate for any sequence of coefficients. In keeping with the custom among functional analysts, I shall refer to the inequality as Bernstein's inequality. (It is not related to the Bernstein inequality for polynomials on the circle.) However, probabilists know the result as Hoeffding's inequality, and the earliest reference known to me is [Hoeffding 1963]. A stronger and more general result goes by the name of the Azuma–Hoeffding inequality; see [Williams 1991], for example.

THEOREM 7.1 (BERNSTEIN'S INEQUALITY). *If $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent Bernoulli random variables and if $a_1, a_2, \dots, a_n$ satisfy $\sum {a_i}^2 = 1$, then for each positive t we have*

$$\text{Prob}\left(\left| \sum_{i=1}^n a_i \varepsilon_i \right| > t \right) \le 2e^{-t^2/2}.$$

This estimate compares well with the probability of finding a standard Gaussian outside the interval $[-t,t]$,

$$\frac{2}{\sqrt{2\pi}}\int_t^\infty e^{-s^2/2}\,ds.$$

The method by which Bernstein's inequality is proved has become an industry standard.

Proof. We start by showing that, for each real λ,

$$Ee^{\lambda\sum a_i\varepsilon_i} \le e^{\lambda^2/2}. \tag{7.1}$$

The idea will then be that $\sum a_i\varepsilon_i$ cannot be large too often, since, whenever it is large, its exponential is enormous.

To prove (7.1), we write

$$Ee^{\lambda\sum a_i\varepsilon_i} = E\prod_1^n e^{\lambda a_i\varepsilon_i}$$

and use independence to deduce that this equals

$$\prod_1^n Ee^{\lambda a_i\varepsilon_i}.$$

For each i the expectation is

$$Ee^{\lambda a_i\varepsilon_i} = \frac{e^{\lambda a_i}+e^{-\lambda a_i}}{2} = \cosh\lambda a_i.$$

Now, $\cosh x \le e^{x^2/2}$ for any real x, so, for each i,

$$Ee^{\lambda a_i\varepsilon_i} \le e^{\lambda^2 a_i^2/2}.$$

Hence

$$Ee^{\lambda\sum a_i\varepsilon_i} \le \prod_1^n e^{\lambda^2 a_i^2/2} = e^{\lambda^2/2},$$

since $\sum a_i^2 = 1$.

To pass from (7.1) to a probability estimate, we use the inequality variously known as Markov's or Chebyshev's inequality: if X is a nonnegative random variable and R is positive, then

$$R\ \mathrm{Prob}(X\ge R)\ \le\ EX$$

(because the integral includes a bit where a function whose value is at least R is integrated over a set of measure $\mathrm{Prob}(X\ge R)$).

Suppose $t\ge 0$. Whenever $\sum a_i\varepsilon_i \ge t$, we will have $e^{t\sum a_i\varepsilon_i}\ge e^{t^2}$. Hence

$$e^{t^2}\,\mathrm{Prob}\left(\sum a_i\varepsilon_i \ge t\right) \le Ee^{t\sum a_i\varepsilon_i} \le e^{t^2/2}$$

by (7.1). So

$$\operatorname{Prob}\left(\sum a_i \varepsilon_i \geq t\right) \;\leq\; e^{-t^2/2},$$

and in a similar way we get

$$\operatorname{Prob}\left(\sum a_i \varepsilon_i \leq -t\right) \;\leq\; e^{-t^2/2}.$$

Putting these inequalities together we get

$$\operatorname{Prob}\left(\left|\sum a_i \varepsilon_i\right| \geq t\right) \;\leq\; 2e^{-t^2/2}. \qquad \square$$

In the next lecture we shall see that deviation estimates that look like Bernstein's inequality hold for a wide class of functions in several geometric settings. For the moment let's just remark that an estimate similar to Bernstein's inequality,

$$\operatorname{Prob}\left(\left|\sum a_i X_i\right| \geq t\right) \;\leq\; 2e^{-6t^2}, \tag{7.2}$$

holds for $\sum a_i^2 = 1$, if the ± 1 valued random variables ε_i are replaced by independent random variables X_i each of which is uniformly distributed on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$. This already has a more geometric flavour, since for these (X_i) the vector $(X_1, X_2, \ldots, X_n)$ is distributed according to Lebesgue measure on the cube $\left[-\frac{1}{2}, \frac{1}{2}\right]^n \subset \mathbb{R}^n$. If $\sum a_i^2 = 1$, then $\sum a_i X_i$ is the distance of the point $(X_1, X_2, \ldots, X_n)$ from the subspace of $\mathbb{R}^n$ orthogonal to $(a_1, a_2, \ldots, a_n)$. So (7.2) says that most of the mass of the cube lies close to any subspace of $\mathbb{R}^n$, which is reminiscent of the situation for the Euclidean ball described in Lecture 1.

Lecture 8. Concentration of Measure in Geometry

The aim of this lecture is to describe geometric analogues of Bernstein's deviation inequality. These geometric deviation estimates are closely related to isoperimetric inequalities. The phenomenon of which they form a part was introduced into the field by V. Milman: its development, especially by Milman himself, led to a new, probabilistic, understanding of the structure of convex bodies in high dimensions. The phenomenon was aptly named the concentration of measure.

We explained in Lecture 5 how the Brunn–Minkowski inequality implies the classical isoperimetric inequality in $\mathbb{R}^n$: among bodies of a given volume, the Euclidean balls have least surface area. There are many other situations where isoperimetric inequalities are known; two of them will be described below. First let's recall that the argument from the Brunn–Minkowski inequality shows more than the isoperimetric inequality.

Let A be a compact subset of $\mathbb{R}^n$. For each point x of $\mathbb{R}^n$, let $d(x, A)$ be the distance from x to A:

$$d(x, A) = \min\left\{|x - y| : y \in A\right\}.$$

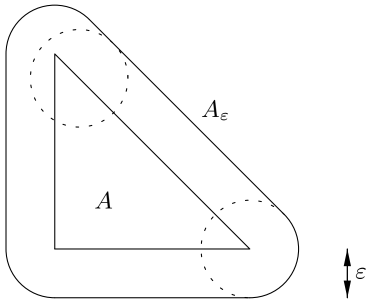

Figure 25. An ε-neighbourhood.

For each positive ε, the Minkowski sum $A + \varepsilon B_2^n$ is exactly the set of points whose distance from A is at most ε. Let's denote such an ε-neighbourhood A_ε; see Figure 25.

The Brunn–Minkowski inequality shows that, if B is an Euclidean ball of the same volume as A, we have

$$\operatorname{vol}(A_\varepsilon) \geq \operatorname{vol}(B_\varepsilon) \quad \text{for any } \varepsilon > 0.$$

This formulation of the isoperimetric inequality makes much clearer the fact that it relates the measure and the metric on $\mathbb{R}^n$. If we blow up a set in $\mathbb{R}^n$ using the metric, we increase the measure by at least as much as we would for a ball.

This idea of comparing the volumes of a set and its neighbourhoods makes sense in any space that has both a measure and a metric, regardless of whether there is an analogue of Minkowski addition. For any metric space (Ω, d) equipped with a Borel measure μ, and any positive α and ε, it makes sense to ask: For which sets A of measure α do the blow-ups A_ε have smallest measure? This general isoperimetric problem has been solved in a variety of different situations. We shall consider two closely related geometric examples. In each case the measure μ will be a probability measure: as we shall see, in this case, isoperimetric inequalities may have totally unexpected consequences.

In the first example, Ω will be the sphere S^{n-1} in $\mathbb{R}^n$, equipped with either the geodesic distance or, more simply, the Euclidean distance inherited from $\mathbb{R}^n$ as shown in Figure 26. (This is also called the *chordal metric;* it was used in Lecture 2 when we discussed spherical caps of given radii.) The measure will be $\sigma = \sigma_{n-1}$, the rotation-invariant probability on S^{n-1}. The solutions of the isoperimetric problem on the sphere are known exactly: they are spherical caps (Figure 26, right) or, equivalently, they are balls in the metric on S^{n-1}. Thus, if a subset A of the sphere has the same measure as a cap of radius r, its neighbourhood A_ε has measure at least that of a cap of radius $r + \varepsilon$.

This statement is a good deal more difficult to prove than the classical isoperimetric inequality on $\mathbb{R}^n$: it was discovered by P. Lévy, quite some time after the isoperimetric inequality in $\mathbb{R}^n$. At first sight, the statement looks innocuous

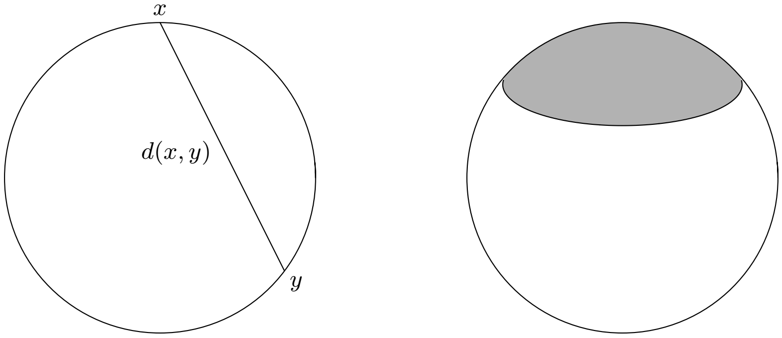

Figure 26. The Euclidean metric on the sphere. A spherical cap (right) is a ball for this metric.

enough (despite its difficulty): but it has a startling consequence. Suppose $\alpha = \frac{1}{2}$, so that A has the measure of a hemisphere H. Then, for each positive ε, the set A_ε has measure at least that of the set H_ε, illustrated in Figure 27. The complement of H_ε is a spherical cap that, as we saw in Lecture 2, has measure about $e^{-n\varepsilon^2/2}$. Hence $\sigma(A_\varepsilon) \geq 1 - e^{-n\varepsilon^2/2}$, so almost the entire sphere lies within distance ε of A, even though there may be points rather far from A. The measure and the metric on the sphere "don't match": the mass of σ concentrates very close to any set of measure $\frac{1}{2}$. This is clearly related to the situation described in Lecture 1, in which we found most of the mass of the ball concentrated near each hyperplane: but now the phenomenon occurs for any set of measure $\frac{1}{2}$.

The phenomenon just described becomes even more striking when reinterpreted in terms of Lipschitz functions. Suppose $f : S^{n-1} \to \mathbb{R}$ is a function on the sphere that is 1-Lipschitz: that is, for any pair of points θ and ϕ on the sphere,

$$|f(\theta) - f(\phi)| \leq |\theta - \phi| .$$

There is at least one number M, the median of f, for which both the sets $(f \leq M)$ and $(f \geq M)$ have measure at least $\frac{1}{2}$. If a point x has distance at most ε from

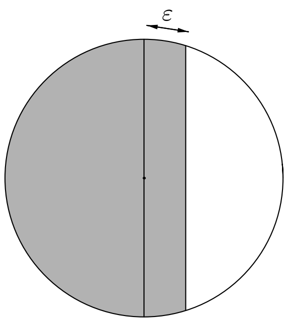

Figure 27. An ε-neighbourhood of a hemisphere.

$(f \le M)$, then (since f is 1-Lipschitz)

$$f(x) \le M + \varepsilon.$$

By the isoperimetric inequality all but a tiny fraction of the points on the sphere have this property:

$$\sigma(f > M + \varepsilon) \ \le \ e^{-n\varepsilon^2/2}.$$

Similarly, f is larger than $M - \varepsilon$ on all but a fraction of the sphere. Putting these statements together we get

$$\sigma(|f - M| > \varepsilon) \ \le \ 2e^{-n\varepsilon^2/2}.$$

So, although f may vary by as much as 2 between a point of the sphere and its opposite, the function is nearly equal to M on almost the entire sphere: f is practically constant.

In the case of the sphere we thus have the following pair of properties.

(i) If $A \subset \Omega$ with $\mu(A) = \frac{1}{2}$ then $\mu(A_\varepsilon) \ge 1 - e^{-n\varepsilon^2/2}$.
(ii) If $f : \Omega \to \mathbb{R}$ is 1-Lipschitz there is a number M for which

$$\mu(|f - M| > \varepsilon) \ \le \ 2e^{-n\varepsilon^2/2}.$$

Each of these statements may be called an approximate isoperimetric inequality. We have seen how the second can be deduced from the first. The reverse implication also holds (apart from the precise constants involved). (To see why, apply the second property to the function given by $f(x) = d(x, A)$.)

In many applications, exact solutions of the isoperimetric problem are not as important as deviation estimates of the kind we are discussing. In some cases where the exact solutions are known, the two properties above are a good deal easier to prove than the solutions: and in a great many situations, an exact isoperimetric inequality is not known, but the two properties are. The formal similarity between property 2 and Bernstein's inequality of the last lecture is readily apparent. There are ways to make this similarity much more than merely formal: there are deviation inequalities that have implications for Lipschitz functions and imply Bernstein's inequality, but we shall not go into them here.

In our second example, the space Ω will be $\mathbb{R}^n$ equipped with the ordinary Euclidean distance. The measure will be the standard Gaussian probability measure on $\mathbb{R}^n$ with density

$$\gamma(x) = (2\pi)^{-n/2} e^{-|x|^2/2}.$$

The solutions of the isoperimetric problem in Gauss space were found by Borell [1975]. They are half-spaces. So, in particular, if $A \subset \mathbb{R}^n$ and $\mu(A) = \frac{1}{2}$, then $\mu(A_\varepsilon)$ is at least as large as $\mu(H_\varepsilon)$, where H is the half-space $\{x \in \mathbb{R}^n : x_1 \le 0\}$ and so $H_\varepsilon = \{x : x_1 \le \varepsilon\}$: see Figure 28.

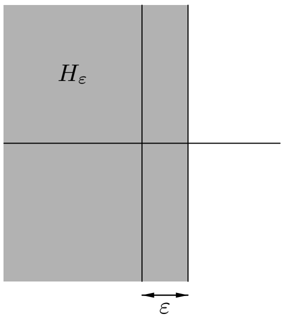

Figure 28. An ε-neighbourhood of a half-space.

The complement of H_ε has measure

$$\frac{1}{\sqrt{2\pi}}\int_\varepsilon^\infty e^{-t^2/2}\,dt \le e^{-\varepsilon^2/2}.$$

Hence,

$$\mu(A_\varepsilon) \ge 1 - e^{-\varepsilon^2/2}.$$

Since n does not appear in the exponent, this looks much weaker than the statement for the sphere, but we shall see that the two are more or less equivalent.

Borell proved his inequality by using the isoperimetric inequality on the sphere. A more direct proof of a deviation estimate like the one just derived was found by Maurey and Pisier, and their argument gives a slightly stronger, Sobolev-type inequality [Pisier 1989, Chapter 4]. We too shall aim directly for a deviation estimate, but a little background to the proof may be useful.

There was an enormous growth in understanding of approximate isoperimetric inequalities during the late 1980s, associated most especially with the name of Talagrand. The reader whose interest has been piqued should certainly look at Talagrand's gorgeous articles [1988; 1991a], in which he describes an approach to deviation inequalities in product spaces that involves astonishingly few structural hypotheses. In a somewhat different vein (but prompted by his earlier work), Talagrand [1991b] also found a general principle, strengthening the approximate isoperimetric inequality in Gauss space. A simplification of this argument was found by Maurey [1991]. The upshot is that a deviation inequality for Gauss space can be proved with an extremely short argument that fits naturally into these notes.

THEOREM 8.1 (APPROXIMATE ISOPERIMETRIC INEQUALITY FOR GAUSS SPACE). *Let $A \subset \mathbb{R}^n$ be measurable and let μ be the standard Gaussian measure on $\mathbb{R}^n$. Then*

$$\int e^{d(x,A)^2/4}\,d\mu \le \frac{1}{\mu(A)}.$$

Consequently, if $\mu(A) = \frac{1}{2}$,

$$\mu(A_\varepsilon) \;\ge\; 1 - 2e^{-\varepsilon^2/4}.$$

PROOF. We shall deduce the first assertion directly from the Prékopa–Leindler inequality (with $\lambda = \frac{1}{2}$) of Lecture 5. To this end, define functions f, g, and m on $\mathbb{R}^n$, as follows:

$$\begin{aligned} f(x) &= e^{d(x,A)^2/4}\,\gamma(x), \\ g(x) &= \chi_A(x)\,\gamma(x), \\ m(x) &= \gamma(x), \end{aligned}$$

where γ is the Gaussian density. The assertion to be proved is that

$$\left(\int e^{d(x,A)^2/4}\,d\mu\right)\mu(A) \le 1,$$

which translates directly into the inequality

$$\left(\int_{\mathbb{R}^n} f\right)\left(\int_{\mathbb{R}^n} g\right) \le \left(\int_{\mathbb{R}^n} m\right)^2.$$

By the Prékopa–Leindler inequality it is enough to check that, for any x and y in $\mathbb{R}^n$,

$$f(x)g(y) \le m\left(\frac{x+y}{2}\right)^2.$$

It suffices to check this for $y \in A$, since otherwise $g(y) = 0$. But, in this case, $d(x, A) \le |x - y|$. Hence

$$\begin{aligned} (2\pi)^n f(x)g(y) &= e^{d(x,A)^2/4} e^{-x^2/2} e^{-y^2/2} \\ &\le \exp\left(\frac{|x-y|^2}{4} - \frac{|x|^2}{2} - \frac{|y|^2}{2}\right) = \exp\left(-\frac{|x+y|^2}{4}\right) \\ &= \left(\exp\left(-\frac{1}{2}\left|\frac{x+y}{2}\right|^2\right)\right)^2 = (2\pi)^n\, m\left(\frac{x+y}{2}\right)^2, \end{aligned}$$

which is what we need.

To deduce the second assertion from the first, we use Markov's inequality, very much as in the proof of Bernstein's inequality of the last lecture. If $\mu(A) = \frac{1}{2}$, then

$$\int e^{d(x,A)^2/4}\,d\mu \le 2.$$

The integral is at least

$$e^{\varepsilon^2/4}\mu\big(d(x,A) \ge \varepsilon\big).$$

So

$$\mu\big(d(x,A) \ge \varepsilon\big) \;\le\; 2e^{-\varepsilon^2/4},$$

and the assertion follows. □

It was mentioned earlier that the Gaussian deviation estimate above is essentially equivalent to the concentration of measure on S^{n-1}. This equivalence depends upon the fact that the Gaussian measure in $\mathbb{R}^n$ is concentrated in a spherical shell of thickness approximately 1, and radius approximately $\sqrt{n}$. (Recall that the Euclidean ball of volume 1 has radius approximately $\sqrt{n}$.) This concentration is easily checked by direct computation using integration in spherical polars: but the inequality we just proved will do the job instead. There is an Euclidean ball of some radius R whose Gaussian measure is $\frac{1}{2}$. According to the theorem above, Gaussian measure concentrates near the boundary of this ball. It is not hard to check that R is about $\sqrt{n}$. This makes it quite easy to show that the deviation estimate for Gaussian measure guarantees a deviation estimate on the sphere of radius $\sqrt{n}$ with a decay rate of about $e^{-\varepsilon^2/4}$. If everything is scaled down by a factor of $\sqrt{n}$, onto the sphere of radius 1, we get a deviation estimate that decays like $e^{-n\varepsilon^2/4}$ and n now appears in the exponent. The details are left to the reader.

The reader will probably have noticed that these estimates for Gauss space and the sphere are not quite as strong as those advertised earlier, because in each case the exponent is $\ldots\varepsilon^2/4\ldots$ instead of $\ldots\varepsilon^2/2\ldots$. In some applications, the sharper results are important, but for our purposes the difference will be irrelevant. It was pointed out to me by Talagrand that one can get as close as one wishes to the correct exponent $\ldots\varepsilon^2/2\ldots$ by using the Prékopa–Leindler inequality with λ close to 1 instead of $\frac{1}{2}$ and applying it to slightly different f and g.

For the purposes of the next lecture we shall assume an estimate of $e^{-\varepsilon^2/2}$, even though we proved a weaker estimate.

Lecture 9. Dvoretzky's Theorem

Although this is the ninth lecture, its subject, Dvoretzky's Theorem, was really the start of the modern theory of convex geometry in high dimensions. The phrase "Dvoretzky's Theorem" has become a generic term for statements to the effect that high-dimensional bodies have almost ellipsoidal slices. Dvoretzky's original proof shows that any symmetric convex body in $\mathbb{R}^n$ has almost ellipsoidal sections of dimension about $\sqrt{\log n}$. A few years after the publication of Dvoretzky's work, Milman [Milman 1971] found a very different proof, based upon the concentration of measure, which gave slices of dimension $\log n$. As we saw in Lecture 2 this is the best one can get in general. Milman's argument gives the following.

THEOREM 9.1. *There is a positive number c such that, for every $\varepsilon > 0$ and every natural number n, every symmetric convex body of dimension n has a slice of dimension*

$$k \geq \frac{c\varepsilon^2}{\log(1+\varepsilon^{-1})} \log n$$

that is within distance $1+\varepsilon$ *of the* k*-dimensional Euclidean ball.*

There have been many other proofs of similar results over the years. A particularly elegant one [Gordon 1985] gives the estimate $k \geq c\varepsilon^2 \log n$ (removing the logarithmic factor in ε^{-1}), and this estimate is essentially best possible. We chose to describe Milman's proof because it is conceptually easier to motivate and because the concentration of measure has many other uses. A few years ago, Schechtman found a way to eliminate the log factor within this approach, but we shall not introduce this subtlety here. We shall also not make any effort to be precise about the dependence upon ε.

With the material of Lecture 8 at our disposal, the plan of proof of Theorem 9.1 is easy to describe. We start with a symmetric convex body and we consider a linear image K whose maximal volume ellipsoid is the Euclidean ball. For this K we will try to find almost spherical sections, rather than merely ellipsoidal ones. Let $\|\cdot\|$ be the norm on $\mathbb{R}^n$ whose unit ball is K. We are looking for a k-dimensional space H with the property that the function

$$\theta \mapsto \|\theta\|$$

is almost constant on the Euclidean sphere of H, $H \cap S^{n-1}$. Since K contains B_2^n, we have $\|x\| \leq |x|$ for all $x \in \mathbb{R}^n$, so for any θ and ϕ in S^{n-1},

$$\big|\|\theta\| - \|\phi\|\big| \leq \|\theta - \phi\| \leq |\theta - \phi|.$$

Thus $\|\cdot\|$ is a Lipschitz function on the sphere in $\mathbb{R}^n$, (indeed on all of $\mathbb{R}^n$). (We used the same idea in Lecture 4.) From Lecture 8 we conclude that the value of $\|\theta\|$ is almost constant on a very large proportion of S^{n-1}: it is almost equal to its average

$$M = \int_{S^{n-1}} \|\theta\| \, d\sigma,$$

on most of S^{n-1}.

We now choose our k-dimensional subspace at random. (The exact way to do this will be described below.) We can view this as a random embedding

$$T : \mathbb{R}^k \to \mathbb{R}^n.$$

For any particular unit vector $\psi \in \mathbb{R}^k$, there is a very high probability that its image $T\psi$ will have norm $\|T\psi\|$ close to M. This means that even if we select quite a number of vectors $\psi_1, \psi_2, \ldots, \psi_m$ in S^{k-1} we can guarantee that there will be some choice of T for which *all* the norms $\|T\psi_i\|$ will be close to M. We will thus have managed to pin down the radius of our slice in many different directions. If we are careful to distribute these directions well over the sphere in $\mathbb{R}^k$, we may hope that the radius will be almost constant on the entire sphere. For these purposes, "well distributed" will mean that all points of the sphere in $\mathbb{R}^k$ are close to one of our chosen directions. As in Lecture 2 we say that a set $\{\psi_1, \psi_2, \ldots, \psi_m\}$ in S^{k-1} is a δ*-net* for the sphere if every point of S^{k-1} is within

(Euclidean) distance δ of at least one ψ_i. The arguments in Lecture 2 show that S^{k-1} has a δ-net with no more than

$$m = \left(\frac{4}{\delta}\right)^k$$

elements. The following lemma states that, indeed, pinning down the norm on a very fine net, pins it down everywhere.

Lemma 9.2. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^k$ and suppose that for each point ψ of some δ-net on S^{k-1}, we have*

$$M(1-\gamma) \le \|\psi\| \le M(1+\gamma)$$

for some $\gamma > 0$. Then, for every *$\theta \in S^{k-1}$,*

$$\frac{M(1-\gamma-2\delta)}{1-\delta} \le \|\theta\| \le \frac{M(1+\gamma)}{1-\delta}.$$

Proof. Clearly the value of M plays no real role here so assume it is 1. We start with the upper bound. Let C be the maximum possible ratio $\|x\|/|x|$ for nonzero x and let θ be a point of S^{k-1} with $\|\theta\| = C$. Choose ψ in the δ-net with $|\theta - \psi| \le \delta$. Then $\|\theta-\psi\| \le C|\theta-\psi| \le C\delta$, so

$$C = \|\theta\| \le \|\psi\| + \|\theta - \psi\| \le (1+\gamma) + C\delta.$$

Hence

$$C \le \frac{(1+\gamma)}{1-\delta}.$$

To get the lower bound, pick some θ in the sphere and some ψ in the δ-net with $|\psi - \theta| \le \delta$. Then

$$(1-\gamma) \le \|\psi\| \le \|\theta\| + \|\psi-\theta\| \le \|\theta\| + \frac{(1+\gamma)}{1-\delta}|\psi-\theta| \le \|\theta\| + \frac{(1+\gamma)\delta}{1-\delta}.$$

Hence

$$\|\theta\| \ge \left(1-\gamma-\frac{\delta(1+\gamma)}{1-\delta}\right) = \frac{(1-\gamma-2\delta)}{1-\delta}. \qquad \square$$

According to the lemma, our approach will give us a slice that is within distance

$$\frac{1+\gamma}{1-\gamma-2\delta}$$

of the Euclidean ball (provided we satisfy the hypotheses), and this distance can be made as close as we wish to 1 if γ and δ are small enough.

We are now in a position to prove the basic estimate.

Theorem 9.3. *Let K be a symmetric convex body in $\mathbb{R}^n$ whose ellipsoid of maximal volume is B_2^n and put*

$$M = \int_{S^{n-1}} \|\theta\| \, d\sigma$$

as above. Then K has almost spherical slices whose dimension is of the order of nM^2.

PROOF. Choose γ and δ small enough to give the desired accuracy, in accordance with the lemma.

Since the function $\theta \mapsto \|\theta\|$ is Lipschitz (with constant 1) on S^{n-1}, we know from Lecture 8 that, for any $t \geq 0$,

$$\sigma\big(\big|\|\theta\| - M\big| > t\big) \leq 2e^{-nt^2/2}.$$

In particular,

$$\sigma\big(\big|\|\theta\| - M\big| > M\gamma\big) \leq 2e^{-nM^2\gamma^2/2}.$$

So

$$M(1-\gamma) \leq \|\theta\| \leq M(1+\gamma)$$

on all but a proportion $2e^{-nM^2\gamma^2/2}$ of the sphere.

Let $\mathcal{A}$ be a δ-net on the sphere in $\mathbb{R}^k$ with at most $(4/\delta)^k$ elements. Choose a random embedding of $\mathbb{R}^k$ in $\mathbb{R}^n$: more precisely, fix a particular copy of $\mathbb{R}^k$ in $\mathbb{R}^n$ and consider its images under orthogonal transformations U of $\mathbb{R}^n$ as a random subspace with respect to the invariant probability on the group of orthogonal transformations. For each fixed ψ in the sphere of $\mathbb{R}^k$, its images $U\psi$, are uniformly distributed on the sphere in $\mathbb{R}^n$. So for each $\psi \in \mathcal{A}$, the inequality

$$M(1-\gamma) \leq \|U\psi\| \leq M(1+\gamma)$$

holds for U outside a set of measure at most $2e^{-nM^2\gamma^2/2}$. So there will be at least one U for which this inequality holds for *all* ψ in $\mathcal{A}$, as long as the sum of the probabilities of the bad sets is at most 1. This is guaranteed if

$$\left(\frac{4}{\delta}\right)^k 2e^{-nM^2\gamma^2/2} < 1.$$

This inequality is satisfied by k of the order of

$$nM^2 \frac{\gamma^2}{2\log(4/\delta)}. \qquad \square$$

Theorem 9.3 guarantees the existence of spherical slices of K of large dimension, provided the average

$$M = \int_{S^{n-1}} \|\theta\| \, d\sigma$$

is not too small. Notice that we certainly have $M \leq 1$ since $\|x\| \leq |x|$ for all x. In order to get Theorem 9.1 from Theorem 9.3 we need to get a lower estimate for M of the order of

$$\frac{\sqrt{\log n}}{\sqrt{n}}.$$

This is where we must use the fact that B_2^n is the *maximal* volume ellipsoid in K. We saw in Lecture 3 that in this situation $K \subset \sqrt{n}B_2^n$, so $\|x\| \geq |x|/\sqrt{n}$ for all x, and hence

$$M \geq \frac{1}{\sqrt{n}}.$$

But this estimate is useless, since it would not give slices of dimension bigger than 1. It is vital that we use the more detailed information provided by John's Theorem.

Before we explain how this works, let's look at our favourite examples. For specific norms it is usually much easier to compute the mean M by writing it as an integral with respect to Gaussian measure on $\mathbb{R}^n$. As in Lecture 8 let μ be the standard Gaussian measure on $\mathbb{R}^n$, with density

$$(2\pi)^{-n/2}e^{-|x|^2/2}.$$

By using polar coordinates we can write

$$\int_{S^{n-1}} \|\theta\|\, d\sigma = \frac{\Gamma(n/2)}{\sqrt{2}\Gamma((n+1)/2)} \int_{\mathbb{R}^n} \|x\|\, d\mu(x) > \frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \|x\|\, d\mu(x).$$

The simplest norm for which to calculate is the ℓ_1 norm. Since the body we consider is supposed to have B_2^n as its maximal ellipsoid we must use $\sqrt{n}B_1^n$, for which the corresponding norm is

$$\|x\| = \frac{1}{\sqrt{n}} \sum_1^n |x_i|.$$

Since the integral of this sum is just n times the integral of any one coordinate it is easy to check that

$$\frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \|x\|\, d\mu(x) = \sqrt{\frac{2}{\pi}}.$$

So for the scaled copies of B_1^n, we have M bounded below by a fixed number, and Theorem 9.3 guarantees almost spherical sections of dimension proportional to n. This was first proved, using exactly the method described here, in [Figiel et al. 1977], which had a tremendous influence on subsequent developments. Notice that this result and Kašin's Theorem from Lecture 4 are very much in the same spirit, but neither implies the other. The method used here does not achieve dimensions as high as $n/2$ even if we are prepared to allow quite a large distance from the Euclidean ball. On the other hand, the volume-ratio argument does not give sections that are very close to Euclidean: the volume ratio is the closest one gets this way. Some time after Kašin's article appeared, the gap between these results was completely filled by Garnaev and Gluskin [Garnaev and Gluskin 1984]. An analogous gap in the general setting of Theorem 9.1, namely that the existing proofs could not give a dimension larger than some fixed multiple of $\log n$, was recently filled by Milman and Schechtman.

What about the cube? This body *has* B_2^n as its maximal ellipsoid, so our job is to estimate

$$\frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \max |x_i| \, d\mu(x).$$

At first sight this looks much more complicated than the calculation for B_1^n, since we cannot simplify the integral of a maximum. But, instead of estimating the mean of the function $\max |x_i|$, we can estimate its median (and from Lecture 8 we know that they are not far apart). So let R be the number for which

$$\mu\left(\max |x_i| \leq R\right) = \mu\left(\max |x_i| \geq R\right) = \tfrac{1}{2}.$$

From the second identity we get

$$\frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \max |x_i| \, d\mu(x) \geq \frac{R}{2\sqrt{n}}.$$

We estimate R from the first identity. It says that the cube $[-R, R]^n$ has Gaussian measure $\frac{1}{2}$. But the cube is a "product" so

$$\mu\left([-R, R]^n\right) = \left(\frac{1}{\sqrt{2\pi}} \int_{-R}^{R} e^{-t^2/2} \, dt \right)^n.$$

In order for this to be equal to $\frac{1}{2}$ we need the expression

$$\frac{1}{\sqrt{2\pi}} \int_{-R}^{R} e^{-t^2/2} \, dt$$

to be about $1 - (\log 2)/n$. Since the expression approaches 1 roughly like

$$1 - e^{-R^2/2},$$

we get an estimate for R of the order of $\sqrt{\log n}$. From Theorem 9.3 we then recover the simple result of Lecture 2 that the cube has almost spherical sections of dimension about $\log n$.

There are many other bodies and classes of bodies for which M can be efficiently estimated. For example, the correct order of the largest dimension of Euclidean slice of the ℓ_p^n balls, was also computed in the paper [Figiel et al. 1977] mentioned earlier.

We would like to know that for a *general* body with maximal ellipsoid B_2^n we have

$$\int_{\mathbb{R}^n} \|x\| \, d\mu(x) \geq (\text{constant}) \sqrt{\log n} \tag{9.1}$$

just as we do for the cube. The usual proof of this goes via the Dvoretzky–Rogers Lemma, which can be proved using John's Theorem. This is done for example in [Pisier 1989]. Roughly speaking, the Dvoretzky–Rogers Lemma builds something like a cube around K, at least in a subspace of dimension about $\frac{n}{2}$, to which we then apply the result for the cube. However, I cannot resist mentioning that the methods of Lecture 6, involving sharp convolution inequalities, can be used to show that among all symmetric bodies K with maximal ellipsoid B_2^n the cube is precisely the one for which the integral in (9.1) is smallest. This is proved by showing that for each r, the Gaussian measure of rK is at most that of $[-r,r]^n$. The details are left to the reader.

This last lecture has described work that dates back to the seventies. Although some of the material in earlier lectures is more recent (and some is much older), I have really only scratched the surface of what has been done in the last twenty years. The book of Pisier to which I have referred several times gives a more comprehensive account of many of the developments. I hope that readers of these notes may feel motivated to discover more.

Acknowledgements

I would like to thank Silvio Levy for his help in the preparation of these notes, and one of the workshop participants, John Mount, for proofreading the notes and suggesting several improvements. Finally, a very big thank you to my wife Sachiko Kusukawa for her great patience and constant love.

References

[Ball 1990] K. M. Ball, "Volumes of sections of cubes and related problems", pp. 251–260 in *Geometric aspects of functional analysis* (Israel Seminar, 1987–1988), edited by J. Lindenstrauss and V. D. Milman, Lecture Notes in Math. **1376**, Springer, 1990.

[Ball 1991] K. M. Ball, "Volume ratios and a reverse isoperimetric inequality", *J. London Math. Soc.* **44** (1991), 351–359.

[Beckner 1975] W. Beckner, "Inequalities in Fourier analysis", *Ann. of Math.* **102** (1975), 159–182.

[Borell 1975] C. Borell, "The Brunn–Minkowski inequality in Gauss space", *Inventiones Math.* **30** (1975), 205–216.

[Bourgain and Milman 1987] J. Bourgain and V. Milman, "New volume ratio properties for convex symmetric bodies in $\mathbb{R}^n$", *Invent. Math.* **88** (1987), 319–340.

[Bourgain et al. 1989] J. Bourgain, J. Lindenstrauss, and V. Milman, "Approximation of zonoids by zonotopes", *Acta Math.* **162** (1989), 73–141.

[Brascamp and Lieb 1976a] H. J. Brascamp and E. H. Lieb, "Best constants in Young's inequality, its converse and its generalization to more than three functions", *Advances in Math.* **20** (1976), 151–173.

[Brascamp and Lieb 1976b] H. J. Brascamp and E. H. Lieb, "On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation", *J. Funct. Anal.* **22** (1976), 366–389.

[Brøndsted 1983] A. Brøndsted, *An introduction to convex polytopes*, Graduate Texts in Math. **90**, Springer, New York, 1983.

[Busemann 1949] H. Busemann, “A theorem on convex bodies of the Brunn–Minkowski type”, *Proc. Nat. Acad. Sci. USA* **35** (1949), 27–31.

[Figiel et al. 1977] T. Figiel, J. Lindenstrauss, and V. Milman, “The dimension of almost spherical sections of convex bodies”, *Acta Math.* **139** (1977), 53–94.

[Garnaev and Gluskin 1984] A. Garnaev and E. Gluskin, “The widths of a Euclidean ball”, *Dokl. A. N. USSR* **277** (1984), 1048–1052. In Russian.

[Gordon 1985] Y. Gordon, “Some inequalities for Gaussian processes and applications”, *Israel J. Math.* **50** (1985), 265–289.

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

[John 1948] F. John, “Extremum problems with inequalities as subsidiary conditions”, pp. 187–204 in *Studies and essays presented to R. Courant on his* 60*th birthday* (Jan. 8, 1948), Interscience, New York, 1948.

[Johnson and Schechtman 1982] W. B. Johnson and G. Schechtman, “Embedding ℓ_p^m into ℓ_1^n”, *Acta Math.* **149** (1982), 71–85.

[Kašin 1977] B. S. Kašin, “The widths of certain finite-dimensional sets and classes of smooth functions”, *Izv. Akad. Nauk SSSR Ser. Mat.* **41**:2 (1977), 334–351, 478. In Russian.

[Lieb 1990] E. H. Lieb, “Gaussian kernels have only Gaussian maximizers”, *Invent. Math.* **102** (1990), 179–208.

[Lı͡usternik 1935] L. A. Lı͡usternik, “Die Brunn–Minkowskische Ungleichung für beliebige messbare Mengen”, *C. R. Acad. Sci. URSS* **8** (1935), 55–58.

[Maurey 1991] B. Maurey, “Some deviation inequalities”, *Geom. Funct. Anal.* **1**:2 (1991), 188–197.

[Milman 1971] V. Milman, “A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies”, *Funkcional. Anal. i Priložen* **5** (1971), 28–37. In Russian.

[Milman 1985] V. Milman, “Almost Euclidean quotient spaces of subspaces of finite dimensional normed spaces”, *Proc. Amer. Math. Soc.* **94** (1985), 445–449.

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

[Rogers 1964] C. A. Rogers, *Packing and covering*, Cambridge U. Press, Cambridge, 1964.

[Schneider 1993] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, Encyclopedia of Math. and its Applications **44**, Cambridge U. Press, 1993.

[Szarek 1978] S. J. Szarek, “On Kashin's almost Euclidean orthogonal decomposition of ℓ_1^n”, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **26** (1978), 691–694.

[Talagrand 1988] M. Talagrand, “An isoperimetric inequality on the cube and the Khintchine-Kahane inequalities”, *Proc. Amer. Math. Soc.* **104** (1988), 905–909.

[Talagrand 1990] M. Talagrand, “Embedding subspaces of L_1 into ℓ_1^N”, *Proc. Amer. Math. Soc.* **108** (1990), 363–369.

[Talagrand 1991a] M. Talagrand, “A new isoperimetric inequality”, *Geom. Funct. Anal.* **1**:2 (1991), 211–223.

[Talagrand 1991b] M. Talagrand, “A new isoperimetric inequality and the concentration of measure phenomenon”, pp. 94–124 in *Geometric aspects of functional analysis* (Israel Seminar, 1989–1990), edited by J. Lindenstrauss and V. D. Milman, Lecture Notes in Math. **1469**, Springer, 1991.

[Tomczak-Jaegermann 1988] N. Tomczak-Jaegermann, *Banach–Mazur distances and finite-dimensional operator ideals*, Pitman Monographs and Surveys in Pure and Applied Math. **38**, Longman, Harlow, 1988.

[Williams 1991] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.

Index

Keith Ball
Department of Mathematics
University College
University of London
London
United Kingdom
kmb@math.ucl.ac.uk