High-Dimensional Convex Geometry

Amit Rajaraman

Last updated December 19, 2020

Contents

§1. Introduction

Definition 1.1. A subset S of a Euclidean space is said to be *convex* if for any $u_1, \dots, u_r \in S$ and non-negative $\lambda_1, \dots, \lambda_r$ such that $\lambda_1 + \cdots + \lambda_r = 1$, the *affine combination* $\sum_{i=1}^r \lambda_i u_i$ is in S as well.

We primarily consider convex bodies, that is, compact and convex subsets of Euclidean spaces here. To put it more succinctly, a convex body is something that "behaves a bit like a Euclidean ball".
A few simple examples of convex bodies on $\mathbb{R}^n$ are:

- the cube $[-1, 1]^n$. Here, the ratio of the radii of the circumscribed ball to the inscribed ball is $\sqrt{n}$, so it is not much like a Euclidean ball.
- the *n-dimensional regular solid simplex* which is the convex hull of $n+1$ equally spaced points. Here, the ratio of the radii of the circumscribed ball to the inscribed ball is n. This ratio is "maximal" in some sense.
- the n-dimensional "octahedron" or *cross-polytope* which is the convex hull of the $2n$ points $(\pm 1, 0, \dots, 0)$, $(0, \pm 1, 0, \dots, 0)$, $\dots, (0, 0, \dots, 0, \pm 1)$. Note that this is the unit ball on the ℓ_1 norm on $\mathbb{R}^n$ so we denote it as B_1^n. Here, the ratio of the radii of the circumscribed ball to the inscribed ball is $\sqrt{n}$.

Definition 1.2. A *cone* in $\mathbb{R}^n$ is the convex hull of a single point and a convex body of dimension $n-1$. In $\mathbb{R}^n$, the volume of a cone of "height" h over a base of $(n-1)$-dimensional volume B is Bh/n.

Since B_1^n is made up of 2^n pieces similar to the piece with non-negative coordinates, which is a cone of height 1 with base analogous to the similar piece in $\mathbb{R}^{n-1}$, the volume of the non-negative section is $1/n!$. Therefore, the $\text{vol}(B_1^n) = 2^n/n!$.

1.1. The Euclidean Ball

The fourth and final example is the Euclidean ball itself, namely

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

Let us now attempt to calculate $v_n = \text{vol}(B_2^n)$. Note that we can easily get the "surface area" of the ball from the volume by splitting it into "thin" cones from 0 and observing that the volume of each cone is equal to $1/n$ times its base area. Therefore, the surface area of the ball is nv_n.
We perform integration in spherical polar coordinates using two variables - r, which denotes the distance from 0 and θ, which is a point on the unit ball that represents the direction of the point. We obviously have $x = r\theta$. The point θ carries the information of $n-1$ coordinates.
We can then write the integral of a general function on $\mathbb{R}^n$ by

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

Here, $\mathrm{d}\theta$ represents the area measure on the sphere. From our earlier observation, its total mass is nv_n. The r^{n-1} factor appears because the sphere of radius r has r^{n-1} times that of the sphere of radius 1.
An important thing to note about the measure corresponding to $\mathrm{d}\theta$ is that it is *rotation-invariant.* If A is a subset of the sphere and U is orthogonal to A, then UA has the same measure as A. Therefore, we often simplify integrals such as 1.1 by pulling out the nv_n factor to get

$$\int_{\mathbb{R}^n} f = nv_n \int_{r=0}^{\infty} \int_{S^{n-1}} f(r\theta) r^{n-1} \,\mathrm{d}\sigma_{n-1}(\theta) \,\mathrm{d}r \tag{1.2}$$

where σ_{n-1} is the rotation-invariant measure on $\mathbb{R}^{n-1}$ of total mass 1. Now, to evaluate v_n, we choose a suitable f such that the integrals on either side can easily be calculated, namely

$$f : x \mapsto \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right).$$

Then the integral on the left of 1.2 is

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} \prod_{i=1}^{n} \exp\left(-\frac{x_i^2}{2}\right) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \exp\left(-\frac{x_i^2}{2}\right) = \left(\sqrt{2\pi}\right)^n$$

and the integral on the right is

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

Equating the two,

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

Using Stirling's Formula, we can approximate this slightly better as

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

This is quite small for large n. The radius of a ball of volume 1 would be approximately $\sqrt{n/2\pi e}$, which is very large!

This is possibly the first hint one should take that following your intuition is probably not a good idea when dealing with high dimensional spaces.

Let us now restrict ourselves to considering the ball of volume 1.
What is the $(n-1)$-dimensional volume of a slice through the center of the ball? Since the slice is an $(n-1)$ dimensional ball, it is equal to

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

This is approximately equal to $\sqrt{e}$ (using Stirling's formula once again). More generally, the volume of the slice that is at distance x from the center of the ball is equal to

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

Note that this is normally distributed but the variance $1/2\pi e$ *does not depend on* n![1] So despite the fact that the radius grows as $\sqrt{n}$, the distribution of the volume stays the same. For example, nearly all the volume (around 96%) is concentrated in the slab with $\|x_1\| \leq 1/2$.

This might lead us to believe that since the volume is concentrated around any such equator[2] around a subspace, the volume should be concentrated around the intersection of all such equators, which seems to suggest that it should be concentrated around the center. However, for large n, we obviously know that most of the volume should be concentrated on the surface of the sphere[3]. These two points seem to be directly contradictory! However, as might

[1] exclamation mark, not factorial!
[2] since we could have equally well taken something other than x_1.
[3] the volume of a ball of radius dr $(d<1)$ is $d^n \ll 1$ times that of a ball of radius r.

be expected, this is once again because our intuition fails when dealing with high-dimensional spaces.
The measure of unit ball is "concentrated" both near the surface *and* around the equator, for any equator. To make more sense of this[4], while each x_i is small, the overall distance from 0 is quite large since the small individual coordinates are compensated by the large dimension n. The former leads to the point being close to the equator and the latter leads to the point being close to the surface of the ball.

Another fun[5] thing to think about is the following. Consider the cube $[-1,1]^n$. Construct a ball of radius 1/2 at each of the 2^n vertices $(\pm 1, \ldots, \pm 1)$. Now, construct the ball with center $(\frac{1}{2}, \ldots, \frac{1}{2})$ that touches each of these 2^n balls. Then note that for $n = 4$, this ball touches (the center of each face of) the cube, and for $n \geq 5$, it actually goes *outside* the cube!

To conclude, let us write the volume of a general convex body K in spherical polar coordinates. Assume that K has 0 in its interior and for each direction $\theta \in S^{n-1}$, let $r(\theta)$ be the radius of K (in that direction). Then,

$$\operatorname{vol}(K) = n v_n \int_{S^{n-1}} \int_0^{r(\theta)} s^{n-1} \, \mathrm{d}s \, \mathrm{d}\sigma = v_n \int_{S^{n-1}} r(\theta)^n \, \mathrm{d}\sigma.$$

Definition 1.3. A convex body K is said to be *(centrally) symmetric* if $-x \in K$ whenever $x \in K$.

Any symmetric body (other than the trivial $\{0\}$) is the unit ball under some $\|\cdot\|_K$ on $\mathbb{R}^n$ (for example, the octahedron was the unit ball under the ℓ_1 norm). For a general symmetric body K, the volume is given by

$$\operatorname{vol}(K) = v_n \int_{S^{n-1}} \|\theta\|_K^{-n} \, \mathrm{d}\sigma_{n-1}(\theta) \tag{1.3}$$

1.2. The Cube

So for example. since the volume of the cube $[-1,1]^n$ is 2^n, we can use it to estimate the average radius of the cube as

$$v_n \int_{S^{n-1}} r(\theta)^n = 2^n \implies \int_{S^{n-1}} r(\theta)^n \approx \left(\sqrt{\frac{2n}{\pi e}}\right)^n \text{ so the average radius is approximately } \sqrt{\frac{2n}{\pi e}}$$

That is, the volume of the cube is far more concentrated towards the corners (where the radius is closer to $\sqrt{n}$), rather than the middles of facets (where the radius is closer to 1).
It can actually be shown[6] that the fraction of volume of the intersection of the cube and the ball is less than $\exp(-4n/45)$, which further emphasizes the point that nearly all the volume lies in the corners.

Definition 1.4. A body which is bounded by a finite number of flat facets is called a *polytope*.

A polytope is essentially the intersection of a finite number of half-spaces Note that the cube is a polytope with $2n$ facets.

Earlier, we remarked that the cube is not much like a Euclidean ball. So a question that might come to mind is: If K is a polytope with m facets, how close can K be to the Euclidean ball?

Let us define this "closeness" more concretely.

Definition 1.5 (Banach-Mazur Distance)**.** The *Banach-Mazur 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 $\tilde{L} \subseteq K \subseteq d\tilde{L}$.

[4] The answers to this mathoverflow question might further aid understanding
[5] subject to debate
[6] Consider the random variable $z_i = x_i^2$ where x_i is drawn uniformly randomly from $[-1,1]$. Show that $\mathbf{E}[z_i] = 1/3$ and $\mathbf{Var}[z_i] = 4/45$ and use the Chernoff bound to get a bound on $\Pr[\sum_i z_i \leq 1]$.

Henceforth, we refer to the Banach-Mazur distance as just *distance*.
This corresponds to the how we thought of inscribing/circumscribing a ball earlier, since the ratio of the two radii we considered is just this distance.
If we wanted to make this distance a metric, then we should consider $\log d$ instead of d (the current distance is multiplicative and for any K, $d(K,K) = 1$).

From what we mentioned earlier, we know that the distance between the cube and the Euclidean ball in $\mathbb{R}^n$ is at most $\sqrt{n}$. We shall prove later that it is indeed equal to $\sqrt{n}$.

As might be expected, if we want a polytope that approximates the ball very well, we would need a very large number of facets.

Definition 1.6. For a fixed unit vector v and some $\varepsilon \in [0,1)$, the set

$$C(\varepsilon, v) = \{\theta \in S^{n-1} : \langle \theta, v\rangle \geq \varepsilon\}$$

is called the *ε-cap about v* or more generally, a *spherical cap* (or just *cap*).
It is often better to write a cap in terms of its radius rather than in terms of ε. The *cap of radius r about v* is

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

It is easy to see that a cap of radius r is a $(1 - \frac{r^2}{2})$-cap.
As we shall see in the proof of Theorem 1.3, it is useful to know some upper and lower bounds on the area of an ε-cap.

Lemma 1.1 (Upper bound on the area of spherical caps)**.** For $\varepsilon \in [0,1)$, the cap $C(\varepsilon, u)$ on S^{n-1} has measure (under σ_{n-1}) at most $e^{-n\varepsilon^2/2}$.

Lemma 1.2 (Lower bound on the area of spherical caps)**.** For $r \in [0,2]$, a cap of radius r on S^{n-1} has measure (under σ_{n-1}) at least $\frac{1}{2}(r/2)^{n-1}$.

Proof. Suppose $n \geq 2$ and let $\alpha = 2\sin^{-1}(r/2)$. We can assume that $\alpha \in [0, \frac{\pi}{2}]$ since we can prove the other case similarly. Then the measure of the cap is given by

$$\begin{aligned}
A(n,\alpha) &= \int_0^\alpha \frac{(n-1)v_{n-1}}{nv_n} (\sin\theta)^{n-2}\,\mathrm{d}\theta \\
&= \frac{(n-1)\Gamma\left(\frac{n}{2}+1\right)}{n\Gamma\left(\frac{n-1}{2}+1\right)\sqrt{\pi}} \int_0^\alpha \sin^{n-2}(\theta)\,\mathrm{d}\theta \\
&= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi}} \int_0^\alpha \sin^{n-2}(\theta)\,\mathrm{d}\theta \\
&\geq \frac{1}{\sqrt{\pi}} \int_0^\alpha \left(\frac{2\theta}{\pi}\right)^{n-2} \mathrm{d}\theta \\
&= \frac{1}{\sqrt{\pi}} \cdot \left(\frac{2}{\pi}\right)^{n-2} \frac{\alpha^{n-1}}{n-1} \\
&= \frac{4}{\sqrt{\pi}(n-1)} \left(\frac{4}{\pi}\right)^{n-2} \cdot \frac{1}{2}\,(\alpha/2)^{n-1} \\
&\geq \frac{4}{\sqrt{\pi}(n-1)} \left(\frac{4}{\pi}\right)^{n-2} \cdot \frac{1}{2}(r/2)^{n-1} \\
&= \frac{\sqrt{\pi}}{n-1} \left(\frac{4}{\pi}\right)^{n-1} \cdot \frac{1}{2}(r/2)^{n-1}
\end{aligned}$$

It is easily shown that

$$\frac{\sqrt{\pi}}{n-1}\left(\frac{4}{\pi}\right)^{n-1} \geq 1$$

for all $n \geq 2$, thus proving the inequality. ∎

Theorem 1.3. Let K be a symmetric polytope in $\mathbb{R}^n$ with $d(K, B_2^n) = d$. Then K has at least $\exp(n/2d^2)$ 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.

Before proving the above theorem, let us reformulate what a symmetric polytope is in another way. Suppose you have a symmetric polytope K with m pairs of facets. Then it is basically the intersection of m slabs in $\mathbb{R}^n$ each of the form $\{x : |\langle x, v_i \rangle| \leq 1\}$ for some $v_i \in \mathbb{R}^n$. That is,

$$K = \{x : |\langle x, v_i \rangle| \leq 1 \text{ for } 1 \leq i \leq m\} \tag{1.4}$$

We can then consider a linear map from $K \to \mathbb{R}^m$ given by

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

This maps $\mathbb{R}^n$ to a subspace of $\mathbb{R}^m$. By the formulation of K given in 1.4 , the intersection of this subspace with the unit cube is just the image of K under T! This is just an n-dimensional slice of $[-1, 1]^m$. Even conversely, any n-dimensional slice of $[-1, 1]^m$ is a convex body with at most m pairs of facets.

Proof. For the proof, let us write what it means for each v_i (following the above notation) if $B_2^n \subseteq K \subseteq dB_2^n$.

- The first inclusion just says that each v_i is of length at most 1 (otherwise, one could consider $v_i / \|v_i\|$, which would be in B_2^n but not in K).
- The latter says that if $\|x\| > d$, then there is some i for which $\langle x, v_i \rangle > 1$. That is, for any unit vector θ, there is some i such that
$$\langle \theta, v_i \rangle > \frac{1}{d}.$$

Since we want to minimize m while satisfying the above two conditions, we can clearly do no better than have $\|v_i\| = 1$ for each i. We want that every $\theta \in S^{n-1}$ is in one of the m $(1/d)$-caps about the (v_i).

- Obviously, to do this, we should attempt to estimate the area of a general ε-cap ($\varepsilon = 1/d$ here). Given Lemma 1.1, we get that
$$m \geq \frac{1}{\exp(-n\varepsilon^2/2)} = \exp\left(\frac{n}{2d^2}\right).$$
- To show that there exists a polytope with the given number of facets, it is enough to find $2 \cdot 4^{n-1}$ points $v_1, \ldots, v_m$ such that the caps of radius 1 centered at these points covers the sphere. Such a set is called a 1-*net*. Now, suppose we choose a set of points on the sphere such that any two of them are at least distance 1 apart. Such a set is called a 1-*separated set*.

  Note that the caps of radius 1/2 centered at each of the points in a 1-separated set are disjoint. Since the measure of a cap of radius 1/2 is at least $1/2 \cdot 4^{n-1}$ (by Lemma 1.2), the number of points in a 1-separated set is at most $2 \cdot 4^{n-1}$.
  It is then enough to choose a "maximal" 1-separated set (a 1-separated set S such that $S \cup x$ is not 1-separated for any $x \in S^{n-1}$) since it is then automatically a 1-net!

  Therefore, there is a 1-net with at most $2 \cdot 4^{n-1}$ points.

  The corresponding polytope then has twice as many, that is, 4^n facets (since each v_i corresponds to a pair of facets).

■

1.3. Fritz John's Theorem

At the very beginning, we had mentioned that the distance of the cube $[-1,1]^n$ and the regular solid simplex are at distance at most $\sqrt{n}$ and n from the ball respectively. However, how would one go about proving that the distances are *exactly* $\sqrt{n}$ and n?

Fritz John's Theorem aids us in this pursuit.
He considered ellipsoids inside convex bodies. If (e_i) is an orthonormal basis of $\mathbb{R}^n$ and (α_i) are positive numbers, then the ellipsoid defined by

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

has volume equal to $v_n \prod_i \alpha_i$. The theorem states that there is a *unique* maximal ellipsoid contained in any convex body, and further, he characterized this ellipsoid! Also, if K is a symmetric convex body and $\mathcal{E}$ is its maximal ellipsoid, then $K \subseteq \sqrt{n}\mathcal{E}$!
We can then use this characterization combined with an affine transformation to prove that the distance between the cube and the ball is $\sqrt{n}$.
We state John's Theorem after performing the affine transformation, since it is easier to understand what's going on then. Roughly, it says that there should be several points of contact between the ball and the boundary of K.

Theorem 1.4 (Fritz John's Theorem)**.** Each convex body K contains a unique ellipsoid of maximal volume. This ellipsoid is B_2^n iff $B_2^n \subseteq K$ and for some m, there are unit vectors $(u_i)_1^m$ on the boundary of K and positive numbers $(c_i)_1^m$ such that

$$\sum_i c_i u_i = 0 \tag{1.5}$$

and for each $x \in \mathbb{R}^n$,

$$\sum_i c_i \langle x, u_i\rangle^2 = \|x\|^2. \tag{1.6}$$

Before proving the theorem, let us discuss some of its implications.

The first condition essentially says that the (u_i) are not all on one side of the body[7]. Intuitively, this makes sense because if the points were concentrated towards one side of the body, then we could move the ball a little bit in the opposite direction and then expand it a little.
The second says that the (u_i) are something like an orthonormal basis, in that we can resolve the norm as a weighted sum of squares of inner products.
On the other hand, Equation (1.6) is equivalent to saying that for all $x \in \mathbb{R}^n$,

$$x = \sum_i c_i \langle x, u_i\rangle u_i.$$

This ensures that the points do not lie close to a (proper) subspace of $\mathbb{R}^n$. This makes sense intuitively as well since if they did, we could contract the ellipsoid a bit in this direction and expand it orthogonally.

Equation (1.6) is written more compactly as

$$\sum_i c_i u_i \otimes u_i = I_n. \tag{1.7}$$

Here, $u_i \otimes u_i$ represents the (rank-1) orthogonal projection onto the span of u_i, the map given by $x \mapsto \langle x, u_i\rangle u_i$. Note that this map is just equal to $u_i u_i^\top$. This implies that the trace of this projection is equal to $\|u_i\|^2 = 1$[8]. Equating

[7] more than simple linear independence since the c_i are positive.

[8] We could have also got this more directly by using the fact that the trace of (a matrix in some basis corresponding to) a linear transformation is the sum of its eigenvalues. For an orthogonal projection, this is just equal to the rank of the target space (which is 1 in this case).

the traces of either side of Equation (1.7), we get

$$\sum_i c_i = n. \tag{1.8}$$

Finally, note that if K is a *symmetric* convex body, then the first condition is obsolete since we can just find any (u_i) satisfying the second condition and replacing it with $+u_i$ and $-u_i$ with each having half the original weight.

Let us now consider a couple of examples to better understand the implications of the theorem.

- For the cube $[0,1]^n$, the maximal ellipsoid is B_2^n as one would expect. The points of contact are the standard basis vectors $(e_i)_1^n$ of $\mathbb{R}^n$ with their negatives, and they do indeed satisfy

$$\sum_i e_i \otimes e_i = I_n.$$

- A slightly more nuanced example is that of the regular solid simplex. Unfortunately, there is no simple or standard way to represent the n-dimensional simplex in n dimensions. It is, however, far natural to represent it in $\mathbb{R}^{n+1}$ by considering the convex hull of the $n+1$ standard basis vectors $(e_i)_1^{n+1}$. We also scale it up by a factor of $\sqrt{n(n+1)}$ (so that the ball contained is B_2^n) such that the $n+1$ points (p_i) we take the convex hull of to get the simplex are given by

$$p_i = \sqrt{n(n+1)}e_i.$$

  This simplex can be parametrized as

$$K = \left\{ x \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} x_i = \sqrt{n(n+1)} \text{ and } x_i \geq 0 \text{ for each } i \right\}.$$

  Similar to the cube, the contact points of the maximal ellipsoid are the centers of each of the facets. More precisely, these $n+1$ endpoints are given by

$$u_i = \frac{\sqrt{n(n+1)}}{n} \left(\sum_{j=1}^{n+1} e_j - e_i \right).$$

  Affinely shifting the hyperplane such that it passes through the origin (making $x_0 = \frac{\sqrt{n(n+1)}}{n+1} \sum_i e_i$ the new origin) and setting the constants c_i as $c = \frac{n}{n+1}$ for each i, for any x in the (unshifted) body,

$$\begin{aligned} \sum_i n(n+1)c_i \langle x - x_0, u_i - x_0 \rangle^2 &= c \sum_i \left(\sum_{j=1}^{n+1} \frac{x_j}{n} - \frac{x_i}{n} - \frac{2}{n+1} + \frac{1}{n+1} \right)^2 \quad \left(\langle x, x_0 \rangle = \langle x_0, u_i \rangle = \langle x_0, x_0 \rangle = \frac{1}{n+1} \right) \\ &= n^2 \sum_i \left(\frac{x_i}{n} - \frac{1}{n(n+1)} \right)^2 \\ &= \|x - x_0\|^2. \end{aligned}$$

  It is easily shown that $\sum_i c_i(u_i - x_0) = 0$ and that each $(u_i - x_0)$ is of magnitude unity, thus proving that the ball touching the centers of the facets (which is an affine shift of B_2^n) is the maximal ellipsoid inside the n-dimensional simplex.

Now, let us prove one of the claims that we made at the beginning of the section.

Theorem 1.5. Suppose that K is a symmetric convex body and B_2^n is the maximal ellipsoid contained in K. Then $K \subseteq \sqrt{n}B_2^n$.
Suppose that K is a convex body and B_2^n is the maximal ellipsoid contained in K. Then $K \subseteq nB_2^n$.

Note that while we have stated the above assuming that B_2^n is the maximal ellipsoid, any convex body in general can be brought to this form by performing an affine shift.

Proof.

- Let x be an arbitrary point in the symmetric body K. Our aim is to show that $\|x\| \leq \sqrt{n}$. Let $(u_i)_1^m$ be the points as described in Fritz John's Theorem. We may assume that if u is in this set, then so is $-u$.
  Now, note that for any i, the tangent plane to K at u_i must coincide with the tangent plane to B_2^n at u_i (otherwise, we would get a contradiction to $B_2^n \subseteq K$). Then, since K is convex, any point in the body must be in the half-space defined by this tangent that contains 0 – this means that $\langle x, u_i \rangle \leq 1$ for each i.
  Then, for each i, we have $\langle x, u_i \rangle \leq 1$ *and* $\langle x, -u_i \rangle \leq 1$ (since we've assumed that if u is in the (u_i), then so is $-u$). That is, $|\langle x, u_i \rangle| \leq 1$ for each i.
  Using the above along with Equation (1.6) and Equation (1.8), we now have
  $$\|x\|^2 = \sum_i c_i \langle x, u_i \rangle^2 \leq \sum_i c_i = n,$$
  which is exactly what we set out to prove!

- Let x be an arbitary point in the convex body. From the first part, we already have that $\langle x, u_i \rangle \leq 1$ for each i. We also have $\langle x, u_i \rangle \geq -\|x\|$ (since $\|u_i\| = 1$). Then,
  $$\begin{aligned} 0 &\leq \sum_i c_i \left(1 - \langle x, u_i \rangle\right)\left(\|x\| + \langle x, u_i \rangle\right) \\ \implies \sum_i c_i \langle x, u_i \rangle^2 &\leq \sum_i c_i \|x\| + (1 - \|x\|) \left\langle x, \sum_i c_i u_i \right\rangle \\ \implies \sum_i c_i \langle x, u_i \rangle^2 &\leq \sum_i c_i \|x\| && \left(\text{since } \sum_i c_i u_i = 0\right) \\ \implies \|x\| &\leq n. && \text{(by Equation (1.6) and Equation (1.8))} \end{aligned}$$

■

Let us now prove Fritz John's Theorem.

Lemma 1.6 (Fritz John's Theorem Pt. 1)**.** Let K be a convex body and for some integer m, let there be unit vectors $(u_i)_1^m$ in ∂K and positive reals $(c_i)_1^m$ satisfying Equation (1.5) and Equation (1.6). Then B_2^n is the unique maximal ellipsoid contained in K.

Proof. Let
$$\mathcal{E} = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n \frac{\langle x, e_j \rangle^2}{\alpha_j^2} \leq 1 \right\}$$
be an ellipsoid in K for some orthonormal basis (e_j) and positive (α_j). We must show that

- $\prod_j \alpha_j \leq 1$ (this implies that B_2^n is a maximal ellipsoid) and
- if $\prod_j \alpha_j = 1$, then for every j, $\alpha_j = 1$ (this implies that B_2^n is *the* maximal ellipsoid).

Now, consider the *dual* of $\mathcal{E}$ given by
$$\mathcal{E}^* = \left\{ y \in \mathbb{R}^n : \sum_{i=1}^n \alpha_j^2 \langle x, e_j \rangle^2 \leq 1 \right\}$$
Observe that we can more concisely describe $\mathcal{E}^*$ as $\{y \in \mathbb{R}^n : \langle y, x \rangle \leq 1 \text{ for all } x \in \mathcal{E}\}$ (Why? Try using the Cauchy-Schwarz inequality)[9].

[9]Interested readers can go through this source as well.

Now, note that since $\mathcal{E} \subseteq K$, for any $x \in \mathcal{E}$ and any i, $\langle x, u_i \rangle \leq 1$ (as proved in the first part of Theorem 1.5). This implies that for every i, $u_i \in \mathcal{E}^*$! We then have

$$\begin{aligned}
\sum_j \alpha_j^2 &= \sum_j \alpha_j^2 \left\| e_j \right\|^2 \\
&= \sum_j \alpha_j^2 \sum_i c_i \langle u_i, e_j \rangle^2 \\
&= \sum_i \left(c_i \sum_j \alpha_j^2 \langle u_i, e_j \rangle^2 \right) \\
&\leq \sum_i c_i = n && \text{(since } u_i \in \mathcal{E}^*\text{)}
\end{aligned}$$

Then, using the AM-GM inequality,

$$\prod_j \alpha_j \leq \left(\frac{1}{n} \sum_j \alpha_j^2 \right)^{n/2} \leq 1.$$

This proves the first part. The second part follows directly as well, since if equality holds in the above equation, then every α_i^2 must be the same (the condition for equality to hold in the AM-GM inequality). ∎

This is the easier of the two directions in Fritz John's Theorem. We now prove the harder.

Lemma 1.7 (Separation Theorem)**.** Let X and Y be two disjoint closed convex bodies in $\mathbb{R}^n$ with at least one of them bounded. Then there exists some $v \in \mathbb{R}^n$ such that for all $x \in X$, $\langle x, v \rangle < b$ and for all $y \in Y$, $\langle y, v \rangle > b$.

We leave the proof of the above to the reader.

Lemma 1.8 (Fritz John's Theorem Pt. 2)**.** Let K be a convex body such that B_2^n is a maximal ellipsoid contained in K. Then, for some integer m, there exist unit vectors $(u_i)_1^m$ in ∂K and positive reals $(c_i)_1^m$ satisfying Equation (1.5) and Equation (1.7).

Proof. We want to show that there exist unit vectors (u_i) in ∂K and positive constants (c_i) such that

$$\frac{1}{n} I_n = \sum_i \left(\frac{c_i}{n} \right) (u_i \otimes u_i)$$

Since $\sum_i c_i = n$, we essentially aim to show that $\frac{1}{n} I_n$ is in the convex hull of the $(u_i \otimes u_i)$ (in the space of matrices). To this end, define

$$T = \operatorname{Conv}\left(\{ u \otimes u : u \text{ is a unit vector in } \partial K \} \right).$$

We refer to such u as "contact points". We want to show that $\frac{1}{n} I_n \in T$. Suppose that it is not (we shall finally show that B_2^n is not a maximal ellipsoid). Then Lemma 1.7 implies that there exists a matrix $H = (h_{i,j})$ such that the linear map φ from the set of matrices to $\mathbb{R}$ defined by

$$(a_{i,j}) \mapsto \sum_{i,j} h_{i,j} a_{i,j}$$

satisfies

$$\varphi\left(\frac{I_n}{n} \right) < \varphi(u \otimes u)$$

for all contact points u. Now, since the matrices on either side are symmetric, we may assume that H is symmetric as well (Why?). And since the matrices on either side have trace equal to 1, adding any constant to the diagonal elements of H leaves the inequality unchanged. Therefore, we may suppose that the trace of H is 0. But this just says that $\varphi(I_n) = 0$!

Therefore, we have essentially found a matrix H such that for any contact point u,

$$u^\top H u > 0. \quad \text{(check that } \varphi(u \otimes u) = u^\top H u\text{)}$$

Now, for $\delta > 0$, consider the ellipsoid defined by

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

We claim that $\mathcal{E}_\delta$ is strictly inside K for sufficiently small δ. Note that for each contact point u,

$$u^\top (I_n + \delta H) u = 1 + \delta \left(u^\top H u \right) > 1$$

so no contact point (of B_2^n) is in $\mathcal{E}_\delta$. For each contact point u, consider a neighbourhood n_u such that for all $x \in n_u$, $x^\top (I_n + \delta H) x > 0$ – we know that such a neighbourhood exists due to the continuity of $x \mapsto x^\top (I_n + \delta H) x$. Let N be the union of all these neighbourhoods.
We now want to show that for any $x \in \partial K \setminus N$, $x^\top (I_n + \delta H) x > 1$. To this end, let $\lambda_{\min}$ be the minimum eigenvalue of H. For any $x \in \partial K \setminus N$, $x^\top H x \ge \lambda_{\min} \|x\|^2$. That is, for all such x,

$$x^\top (I_n + \delta H) x \ge (1 + \delta \lambda_{\min}) \|x\|^2.$$

Observe that $\inf_{x \in \partial K \setminus N} \|x\|^2 > 1$.[10] We may also assume that $\lambda_{\min} < 0$, since the claim holds trivially otherwise (we have $\|x\|^2 > 1$). Then, we may set δ as a positive real which is less than $\frac{1}{|\lambda_{\min}|} \left(1 - \frac{1}{\inf_{x \in \partial K \setminus N} \|x\|^2} \right)$. Then for all $x \in \partial K \setminus N$,

$$(1 + \delta \lambda_{\min}) \|x\|^2 > \|x\|^2 \left(1 - \left(1 - \frac{1}{\inf_{y \in \partial K \setminus N} \|y\|^2} \right) \right) \ge 1$$

Therefore, $\mathcal{E}_\delta$ does not intersect ∂K and is *strictly* inside K for sufficiently small δ!

Now, we claim that $\mathcal{E}_\delta$ has volume at least equal to that of B_2^n. Indeed, its volume is given by $v_n / \prod \lambda_i$, where (λ_i) are the eigenvalues of $(I_n + \delta H)$. Since the sum of the eigenvalues is equal to the trace of $I_n + \delta H$, which is n, we can use the AM-GM inequality to get

$$\prod_i \lambda_i \le \left(\frac{1}{n} \sum_i \lambda_i \right)^n = 1,$$

which is exactly what we want, because equality holds iff the eigenvalues are all 1, that is, the ellipsoid is B_2^n (so this leads to a contradiction). ■

Note that we can concatenate the proofs of Lemma 1.8 and Lemma 1.6 to show that *a* maximal ellipsoid is *the* maximal ellipsoid (contained in a convex body).

There is an analogue of Fritz John's Theorem that characterizes the minimal ellipsoid that contains a given body – this is near-direct from the notion of duality that we used in the proof of Lemma 1.6. So for example, it follows form this analogue that the minimal ellipsoid that contains $[-1,1]^n$ is the ball of radius $\sqrt{n}$. This also enables us to say that $d([-1,1]^n, B_2^n)$ is *exactly* equal to $\sqrt{n}$.

There are various extensions of this result. Recall how towards the beginning of these notes we had mentioned how a general convex body K is essentially a unit ball under some norm. Fritz John's Theorem essentially describes linear maps from the Euclidean space to a normed space (under which the unit ball is K) that have largest determinant under the constraing that the Euclidean ball is mapped into K. There is a more general theory that (attempts to) solve this problem under different constraints.

[10] if it was equal to 1, then for any $\varepsilon > 0$, we would be able to find an x such that $\|x\|^2 < 1 + \varepsilon$ (we trivially have that $\|x\|^2 \ge 1$). However, this is not possible because ∂K is compact, we have removed a neighbourhood around each contact point u, and contact points are the only points in ∂K which have norm 1.

§2. Volume Inequalities

2.1. Spherical Sections of Symmetric Bodies

Consider the n-dimensional cross-polytope B_1^n. The maximal ellipsoid in B_1^n is the Euclidean ball of radius $1/\sqrt{n}$. If we take some orthogonal transformation U, then obviously, UB_1^n contains the ball as well, and so does $B_1^n \cap UB_1^n$. But what if we instead consider the minimal ball that contains $B_1^n \cap UB_1^n$? We have the following remarkable theorem, which we prove later.

Theorem 2.1 (Kašin's Theorem)**.** For each n, there is an orthogonal transformation U such that $B_1^n \cap UB_1^n$ is contained in the (Euclidean) ball of radius $32/\sqrt{n}$.

The important thing to note here is the fact that just by intersecting just two copies of the cross-polytope, we manage to reduce the radius of the minimal circumscribing ball by a factor of $\sqrt{n}$! Indeed, this intersection is what we call "approximately spherical" since its distance from the Euclidean ball is then at most 32. The constant factor of 32 can be improved upon as well.

For the same orthogonal transformation U, $\mathrm{Conv}(Q \cup UQ)$ is at distance at most 32 from the Euclidean ball as well (where Q is $[-1,1]^n$).

How would one go about constructing such a transformation? The points of contact between the ball of radius $1/\sqrt{n}$ are those of the form $\left(\pm\frac{1}{n}, \ldots, \pm\frac{1}{n}\right)$.
The points furthest away are those near the corners of the cross-polytope. So we would want to take a transformation whose facets "chop off" these corners.

Recall that in the beginning, we had explained that the volume of the cross-polytope is $2^n/n!$, so if $r(\theta)$ is the radius of B_1^n in the direction θ, then

$$\int_{S^{n-1}} r(\theta)^n \,\mathrm{d}\sigma = \frac{2^n}{n! v_n} \leq \left(\frac{2}{\sqrt{n}}\right)^n. \tag{2.1}$$

This feature wherein $r(\theta)$ is not expected to be much more than $2/\sqrt{n}$ is captured in the following definition.

Definition 2.1 (Volume Ratio)**.** Let K be a convex body in $\mathbb{R}^n$. Then the *volume ratio* of K is defined as

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

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

Equation (2.1) then says that $\mathrm{vr}(B_1^n) \leq 2$ for all n. Let us now prove (a slightly more general version of) Kašin's Theorem, scaling everything up by n for the sake of convenience.

Theorem 2.2. Let K be a symmetric convex body in $\mathbb{R}^n$ that contains B_2^n. Let

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

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

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

Proof. Denote by $\|\cdot\|_K$ the norm under which K is the unit ball. Observe that since $B_2^n \subseteq K$, $\|x\|_K \leq \|x\|$ for all $x \in \mathbb{R}^n$. Note that if U is an orthogonal transformation, then the norm corresponding to $K \cap UK$ is the maximum of that corresponding to K and UK (at that point). Therefore, because the norm corresponding to $8R^2 B_2^n$ is $\frac{1}{8R^2}$ times the Euclidean norm, we just want to find an orthogonal transformation U such that for all $\theta \in S^{n-1}$,

$$\max(\|U\theta\|_K, \|\theta\|_K) \geq \frac{1}{8R^2}.$$

It suffices to show that for all $\theta \in S^{n-1}$,

$$\frac{\|U\theta\|_K + \|\theta\|_K}{2} \geq \frac{1}{8R^2}. \tag{2.2}$$

Now, note that the function N given by $x \mapsto \frac{\|Ux\|_K + \|x\|_K}{2}$ is a norm on $\mathbb{R}^n$. Also, it satisfies $N(x) \leq \|x\|$ for all x. We aim to show that N is "large" everywhere. Let ϕ be a point on the sphere such that $N(\phi) = t$. Then if $\|\theta - \phi\| \leq t$, then

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

That is, for θ in a spherical cap of radius t about ϕ, $N(\theta)$ is at most $2t$. Lemma 1.2 implies that such a cap has measure at least $\frac{1}{2}\left(\frac{t}{2}\right)^{n-1} \geq \left(\frac{t}{2}\right)^n$. Then, considering the integral over only the spherical cap, we have

$$\int_{S^{n-1}} \frac{1}{N(\theta)^{2n}} \,\mathrm{d}\sigma \geq \frac{1}{(2t)^{2n}} \left(\frac{t}{2}\right)^n = \frac{1}{2^{3n} t^n}. \tag{2.3}$$

Now, we claim that there is an orthogonal transformation U such that

$$\int_{S^{n-1}} \frac{1}{N(\theta)^{2n}} \,\mathrm{d}\sigma \leq R^{2n}. \tag{2.4}$$

Because $N(\theta)^2 \geq \|\theta\|_K \|U\theta\|_K$, it suffices to show the existence of an orthogonal transformation U such that

$$\int_{S^{n-1}} \frac{1}{\|\theta\|_K^n \|U\theta\|_K^n} \,\mathrm{d}\sigma \leq R^{2n}.$$

We prove this probabilistically. Consider the average over all orthogonal transformations U of some function f on the sphere. This should just be the average of the value of f over the entire sphere. That is,

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

Setting f as the function given by

$$\theta \mapsto \frac{1}{\|\theta\|_K^n},$$

we have the following:

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

where the last equality follows from Equation (1.3). Since the average of the integral over orthogonal transformations is at most R^{2n}, there must be some orthogonal transformation U such that the integral is at most R^{2n} and Equation (2.4) holds! Then, combining Equation (2.3) and Equation (2.4), we get

$$\frac{1}{2^{3n}t^n} \leq R^{2n} \implies t \geq \frac{1}{8R^2}.$$

That is, for any $\phi \in S^{n-1}$, $\|\phi\|_K \geq \frac{1}{8R^2}$, which is exactly what we set out to show in Equation (2.2)! ∎

Due to the probabilistic nature of the above proof, we do not actually get an orthogonal transformation U. However, a question that might come to mind is - do there exist symmetric bodies for which we *can* explicitly construct U? Consider the simple case of the cross-polytope. As mentioned towards the beginning of this section, we would like to "chop off" the corners. A relatively obvious method to do this that comes to mind is to construct a transformation such that the direction of each of the new corners coincides with the directions of the centers of the original facets. In 2 dimensions, such an orthogonal transformation just means we rotate B_1^2 by $45°$.

However, does such an orthogonal transformation exist for the cross-polytope in any general dimension? Stating it more rigorously, we want to determine for each n if there is an orthogonal transformation U such that for each standard basis vector e_i of $\mathbb{R}^n$, Ue_i is $\sqrt{n}$ times one of the vectors of the form $(\pm\frac{1}{n}, \ldots, \pm\frac{1}{n})$.
That is, we are looking for an $n \times n$ orthogonal matrix with each entry as $\pm\frac{1}{\sqrt{n}}$. Such a matrix without the $\sqrt{n}$ factor (it then merely requires that the rows are orthogonal) is known as a *Hadamard matrix*. For $n \leq 2$, there obviously exist Hadamard matrices

$$\begin{pmatrix} 1 \end{pmatrix} \text{ and } \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}$$

It may be shown that if a Hadamard matrix of dimension $n > 2$ exists, then n is a multiple of 4.[11] However, it is unknown which multiples of 4 Hadamard matrices do indeed exist for. It is known that they do exist for n a power of 2, but even these (known as *Walsh matrices*) don't give good estimates. There are good reasons[12] for believing that we cannot explicitly find an orthogonal transformation that would give the right estimates.

Now, with the aid of Theorem 2.2, proving Theorem 2.1 is near-straightforward.
Recall how in the beginning of this section, we had stated that for the same orthogonal transformation U that is mentioned in Theorem 2.2, $\text{Conv}(Q \cup UQ)$ is at distance 32 from the Euclidean ball, where $Q = [-1,1]^n$. However, we cannot get an approximately spherical body by taking the intersection as we have in Theorem 2.2.
Dually, we cannot get an approximately spherical body by taking the convex hull of a union for a cross-polytope. Both of these ideas (of taking the convex hull *and* the intersection) are combined in the following fascinating result of Milman's.

Theorem 2.3 (QS-Theorem)**.** There is a constant M such that for all symmetric convex bodies K (of any dimension), there are linear maps Q and S and an ellipsoid $\mathcal{E}$ such that if $\tilde{K} = \text{Conv}(K \cup QK)$, then

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

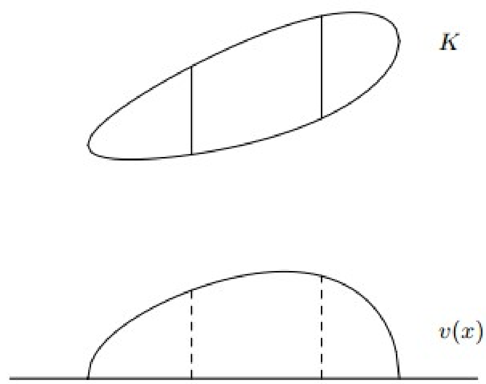
Note that M is a universal constant independent of *everything*. Here the "QS" means "quotient of a subspace".

[11] Let H be a Hadamard matrix. We may assume that all the elements in the first row are $+1$. Let a, b, c and d be the number of columns starting with $(+,+,+)$, $(+,+,-)$, $(+,-,+)$ and $(+,-,-)$ respectively. We trivially have $a+b+c+d=n$. Using the pairwise orthogonality of the first 3 rows, we get 3 other conditions which enable us to conclude that $n = 4a$.

[12] see Ramsey Theory.

2.2. The Prékopa-Leindler Inequality

Consider some convex body K in $\mathbb{R}^2$ and the map $v : \mathbb{R} \to \mathbb{R}$ such that r maps to the length (the Lebesgue measure in $\mathbb{R}$) of the intersection of the line $x = r$ with the body K. We can think of this as "collapsing" the body onto the x-axis like a deck of cards. To understand this better, consider the following image which represents the graph of v for K.[13]



It may be shown that for any convex body K in $\mathbb{R}^2$, the corresponding function v is concave on its support.

How would one go about generalizing this v to a higher dimensional K, say in 3 dimensions? As might be expected, the function $v : \mathbb{R} \to \mathbb{R}$ maps r to the Lebesgue measure (in $\mathbb{R}^2$) of the intersection of $x = r$ with the body K.
Does this v need to be concave? No, it does not! Consider a cone - say the one given by

$$\{(x, y, z) \in \mathbb{R}^3 : y^2 + z^2 \leq x^2, x \geq 0\}.$$

Then since the area of the intersection grows as x^2, the function is quite obviously not concave. However, the cone is a "maximal" convex body in some sense, it is just barely convex and the curved surface is composed of lines. One might now note that the function $r \mapsto \sqrt{v(r)}$ for the cone is indeed (barely) concave! Brunn perhaps noticed this pattern and proved an analogous result for higher dimensions.

Theorem 2.4 (Brunn's Theorem)**.** Let K be a convex body in $\mathbb{R}^n$, u a unit vector in $\mathbb{R}^n$, and for each r, define

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

Then, the function

$$v : r \mapsto \operatorname{vol}(H_r \cap K)^{1/(n-1)}$$

is concave on its support.

A consequence of this theorem is that given any centrally symmetric body in $\mathbb{R}^n$, the $(n-1)$-dimensional slice with the largest area orthogonal to some fixed unit vector u is that through the origin!
Brunn's Theorem was turned from an idle observation to an extremely powerful tool by Minkowski in the form of the Brunn-Minkowski inequality. We omit the proof of Brunn's Theorem as it is obvious from this inequality, which we state shortly.

Before we do this, let us introduce some notation. If X and Y are sets in $\mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, then we write

$$\alpha X + \beta Y = \{\alpha x + \beta y : x \in X, y \in Y\}.$$

[13]Source: An Introduction to Modern Convex Geometry by Keith Ball.

This method of using addition in $\mathbb{R}^n$ to define the addition of sets in $\mathbb{R}^n$ is known as *Minkowski addition*.
In the context of Brunn's Theorem, consider three parallel slices A_r,A_s, and A_t of a body K in $\mathbb{R}^n$ at positions r, s, and t. These slices can be thought of as subsets of $\mathbb{R}^{n-1}$. Further suppose that $r < s < t$ and we have $\lambda \in (0,1)$ such that $s = \lambda r + (1-\lambda)t$. Note that due to the convexity of K,

$$A_s \supseteq \lambda A_r + (1-\lambda)A_t.$$

All Brunn's Theorem says is that

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

Observe that we have removed any remnant of $\mathbb{R}^n$ from this equation. Cleaning it up and restating it more generally, we have the following.

Theorem 2.5 (Brunn-Minkowski Inequality)**.** Let A and B be two non-empty compact subsets of $\mathbb{R}^n$. Then for any $\lambda \in [0,1]$,

$$\operatorname{vol}(\lambda A + (1-\lambda)B)^{1/n} \geq \lambda \operatorname{vol}(A)^{1/n} + (1-\lambda)\operatorname{vol}(B)^{1/n}. \tag{2.5}$$

It is quite obvious that given the above inequality, Brunn's Theorem is true. Here, the non-emptiness of A and B correspond to the fact that we restrict v to the support in Brunn's Theorem.

We encourage the reader to show that Equation (2.5) is equivalent to

$$\operatorname{vol}(A+B)^{1/n} \geq \operatorname{vol}(A)^{1/n} + \operatorname{vol}(B)^{1/n} \tag{2.6}$$

We omit the proof of the Brunn-Minkowski inequality and instead show how it follows from the far more powerful, near-magical Prékopa-Leindler inequality.
Before we do this, let us show how the popular isoperimetric inequality follows from the Brunn-Minkowski inequality.

Theorem 2.6 (Isoperimetric inequality)**.** Among bodies of a given volume, Euclidean balls have the least surface area.

Proof. Let C be a compact body of volume equal to that of B_2^n. The ($(n-1)$-dimensional) "surface area" of C is equal to

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

Equation (2.6) implies that

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

Then,

$$\begin{aligned}\operatorname{vol}(\partial C) &\geq \lim_{\varepsilon \to 0} \frac{n\varepsilon \operatorname{vol}(B_2^n)^{1/n} \operatorname{vol}(C)^{(n-1)/n}}{\varepsilon} \\ &= n \operatorname{vol}(B_2^n) = \operatorname{vol}(\partial B_2^n).\end{aligned}$$

■

It may also be shown using the Brunn-Minkowski Inequality and the weighted AM-GM inequality that for any compact subsets A, B of $\mathbb{R}^n$,

$$\operatorname{vol}(\lambda A + (1-\lambda)B) \geq \operatorname{vol}(A)^{\lambda} \operatorname{vol}(B)^{1-\lambda} \tag{2.7}$$

The above equation is more commonly known as the *multiplicative Brunn-Minkowski inequality*, while Equation (2.5) is known as the *additive Brunn-Minkowski inequality*. It may also be shown that while this is weaker than the Brunn-Minkowski inequality for *particular* subsets A and B, the two are equivalent if we know Equation (2.7) for *all* A and B.

Multiplicative Brunn-Minkowski implies additive Brunn-Minkowski. Fix some $\lambda \in [0,1]$ and let

$$\lambda' = \frac{\dfrac{\lambda}{\operatorname{vol}(B)^{1/n}}}{\dfrac{\lambda}{\operatorname{vol}(B)^{1/n}} + \dfrac{1-\lambda}{\operatorname{vol}(A)^{1/n}}}.$$

Applying Equation (2.7), we get

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

Also,

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

Therefore,

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

which is just additive Brunn-Minkowski. ■

This form is slightly more advantageous because there is no mention of the dimension n or the non-emptiness of A and B.

The Prékopa-Leindler inequality that we mentioned earlier is essentially a generalization of the Brunn-Minkowski inequality to a more functional form, similar to how the Cauchy-Bunyakovasky-Schwarz inequality is a functional analogue of the Cauchy-Schwarz inequality.

To get a little more intuition for how the Brunn-Minkowski inequality is connected to the Prékopa-Leindler inequality, define f as the indicator function on A, g as the indicator function on B, and m as the indicator function on $\lambda A + (1-\lambda)B$, .[14] Then Equation (2.7) says

$$\int_{\mathbb{R}^n} m \geq \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda}.$$

What is the relation between m, f, and g that perhaps leads to this inequality being true? If for some x and y, $f(x) = 1$ and $g(y) = 1$, then we have $m(\lambda x + (1-\lambda)y) = 1$ as well. Therefore, for any $x, y \in \mathbb{R}^n$

$$m(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}.$$

It turns out that this condition is enough to conclude Equation (2.7)!

Theorem 2.7 (Prékopa-Leindler inequality)**.** Let f, g and m be non-negative measurable functions on $\mathbb{R}^n$ and $\lambda \in (0,1)$ such that for all $x, y \in \mathbb{R}^n$,

$$m(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}. \tag{2.8}$$

Then,

$$\int_{\mathbb{R}^n} m \geq \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda}.$$

[14] the indicator function on X for $X \subseteq \mathbb{R}^n$ is the map from $\mathbb{R}^n$ to $\{0,1\}$ such that $f(x) = 1$ if $x \in X$ and 0 otherwise.

The astute reader might notice that this is something of a reversed Hölder's inequality, which says that if we have non-negative functions f and g and define m by $m(z) = f(z)^\lambda g(z)^{1-\lambda}$ for each z, then

$$\int m \leq \left(\int f\right)^\lambda \left(\int g\right)^{1-\lambda}. \tag{2.9}$$

The difference is that in the Prékopa-Leindler inequality, we have

$$m(\lambda x + (1-\lambda)y) \geq \sup_{x,y} f(x)^\lambda g(x)^{1-\lambda},$$

whereas in Hölder's, we only consider the pair $(x, y) = (z, z)$.

Proof of one-dimensional Brunn-Minkowski inequality. Suppose A and B are non-empty measurable subsets of $\mathbb{R}$. We use $\|\cdot\|$ to represent the Lebesgue measure on $\mathbb{R}$.
We can assume that A and B are compact[15]. We can now shift both sets and assume that $A \cap B = \{0\}$. However, in this case, we have $A \cup B \subseteq A + B$ and so, due to the almost-disjointedness of A and B,

$$\|A + B\| \geq \|A \cup B\| = \|A\| + \|B\|.$$

This is just Equation (2.6). ■

Proof of one-dimensional Prékopa-Leindler inequality. We have non-negative measurable functions f, g, and m. We use $\|\cdot\|$ to represent th Lebesgue measure on $\mathbb{R}$. For any function $h : \mathbb{R} \to \mathbb{R}$ and $t \in \mathbb{R}$, define

$$L_h(t) = \{x \in \mathbb{R} : h(x) \geq t\}.$$

Then note that by Equation (2.8),

$$L_m(t) \supseteq \lambda L_f(t) + (1-\lambda) L_g(t).$$

We can then apply the one-dimensional Brunn-Minkowski inequality to get

$$\|L_m(t)\| \geq \|\lambda L_f(t) + (1-\lambda)L_g(t)\| \geq \lambda \|L_f(t)\| + (1-\lambda)\|L_g(t)\|.$$

Finally, we can assume boundedness of all three functions and use Fubini's Theorem to say that

$$\begin{aligned}
\int m &= \int \|L_m(t)\| \,\mathrm{d}t \\
&\geq \lambda \int \|L_f(t)\| \,\mathrm{d}t + (1-\lambda) \int \|L_g(t)\| \,\mathrm{d}t \\
&= \lambda \int f + (1-\lambda) \int g \\
&\geq \left(\int f\right)^\lambda \left(\int g\right)^{1-\lambda},
\end{aligned}$$

where the last step follows from the weighted AM-GM inequality. ■

Proof of Prékopa-Leindler inequality. We prove this inductively. Suppose we have m, f, and g from $\mathbb{R}^n \to \mathbb{R}$ ($n > 1$) satisfying Equation (2.8). For any $z \in \mathbb{R}$ and any function $h : \mathbb{R}^n \to \mathbb{R}$, we denote by $h_z : \mathbb{R}^{n-1} \to \mathbb{R}$ the function given by $h_z(x) = h(x, z)$ (for $z \in \mathbb{R}^{n-1}$) – we make the last coordinate constant and consider the resulting function on the remaining $n-1$ coordinates. Now, let $\alpha, \beta \in \mathbb{R}$, $x, y \in \mathbb{R}^{n-1}$ and let $\gamma = \lambda\alpha + (1-\lambda)\beta$. Then,

$$\begin{aligned}
m_\gamma(\lambda x + (1-\lambda)y) &= m(\lambda x + (1-\lambda)y, \lambda\alpha + (1-\lambda)\beta) \\
&\geq f(x, \alpha)^\lambda g(y, \beta)^{1-\lambda} \\
&= f_\alpha(x)^\lambda g_\beta(y)^{1-\lambda}.
\end{aligned}$$

[15] due to the inner regularity of the Lebesgue measure.

That is, m_γ, f_α, and g_β satisfy Equation (2.8) (on $\mathbb{R}^{n-1}$). We can then apply the inductive hypothesis on them to get

$$\int_{\mathbb{R}^{n-1}} m_\gamma \geq \left(\int_{\mathbb{R}^{n-1}} f_\alpha\right)^\lambda \left(\int_{\mathbb{R}^{n-1}} g_\beta\right)^{1-\lambda}.$$

Now, for any function $h : \mathbb{R}^n \to \mathbb{R}$, we denote by $\tilde{h} : \mathbb{R} \to \mathbb{R}$ the function given by

$$\gamma \mapsto \int_{\mathbb{R}^{n-1}} f_\gamma.$$

Note that the functions $\tilde{m}$, $\tilde{f}$, and $\tilde{g}$ satisfy the condition for the one-dimensional Prékopa-Leindler inequality! Therefore, condensing the iterated integral to a joint integral, we get

$$\int_{\mathbb{R}^n} m \geq \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda},$$

which is exactly what we desire! ■

This proof is quite magical - we use the inequality on $\mathbb{R}^{n-1}$ and $\mathbb{R}^1$ with barely any extra work to conclude that it holds for $\mathbb{R}^n$.
To conclude this section, we state another surprising result in a similar vein to the nice observation that is Brunn's Theorem.

Theorem 2.8 (Busemann's Theorem)**.** 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 convex as well.

2.3. The Reverse Isoperimetric Problem

The Isoperimetric inequality solves the problem of finding the body with the largest volume among bodies with a given surface area. How would one go about solving the reversed problem – finding the body with the largest surface area among bodies with a given volume? We must phrase this more carefully such that it makes sense because as it stands, we could make the surface area arbitrarily large (consider a large thin disc). So the more common way of phrasing it is – given a convex body, how small can we make its surface area by applying an affine transformation that preserves volume?

Theorem 2.9. Let K be a convex body, T a regular solid simplex in $\mathbb{R}^n$, and Q a cube in $\mathbb{R}^n$. Then, there is an affine transformation $\tilde{K}$ of K such that the volume of $\tilde{K}$ is equal to that of T and whose surface area is at most that of T. If K is symmetric, then there is an affine transformation $\tilde{K}$ of K such that the volume of $\tilde{K}$ is equal to that of Q and whose surface area is at most that of Q.

The primary focus of this section is to find the bodies with the largest volume ratios – this is answered for symmetric bodies in Theorem 2.10, which we encourage the reader to look at now.

Given this, we can prove the second part of Theorem 2.9 as follows.
Choose $\tilde{K}$ such that its maximal ellipsoid is B_2^n. Then $\tilde{K}$ has volume at most 2^n (since this is the volume of the cube with maximal ellipsoid B_2^n). Note that

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

Therefore, we shall show that

$$\operatorname{vol}(\partial \tilde{K}) \leq 2n \operatorname{vol}(\tilde{K})^{(n-1)/n}$$

Indeed, we have

$$\begin{aligned}
\operatorname{vol}(\partial\tilde{K}) &= \lim_{\varepsilon\to 0}\frac{\operatorname{vol}(\tilde{K}+\varepsilon B_2^n)-\operatorname{vol}(\tilde{K})}{\varepsilon} \\
&\le \lim_{\varepsilon\to 0}\frac{\operatorname{vol}(\tilde{K}+\varepsilon \tilde{K})-\operatorname{vol}(\tilde{K})}{\varepsilon} && (\text{because } B_2^n\subseteq\tilde{K}) \\
&= \operatorname{vol}(\tilde{K})\lim_{\varepsilon\to 0}\frac{(1+\varepsilon)^n-1}{\varepsilon} \\
&= n\operatorname{vol}(\tilde{K}) \\
&= n\operatorname{vol}(\tilde{K})^{1/n}\operatorname{vol}(\tilde{K})^{(n-1)/n} \\
&\le 2n\operatorname{vol}(\tilde{K})^{(n-1)/n}. && (\text{by Theorem 2.10})
\end{aligned}$$

Theorem 2.10. Among symmetric convex bodies, the cube has the largest volume ratio.

The above is equivalent to saying that if K is a convex body whose maximal ellipsoid is B_2^n, then $\operatorname{vol}(K)\le 2^n$. By Fritz John's Theorem, there exist unit vectors (u_i) and positive (c_i),

$$\sum_i c_i u_i\otimes u_i = I_n.$$

Consider the polytope

$$C=\{x\in\mathbb{R}^n : |\langle x,u_i\rangle|\le 1 \text{ for } 1\le i\le m\}. \tag{2.10}$$

We clearly have $K\subseteq C$, so it suffices to show that $\operatorname{vol}(C)\le 2^n$.
The most important tool we use for this is the following.

Theorem 2.11 (Young's Convolution Inequality)**.** Suppose $f\in L^p(\mathbb{R})$, $g\in L^q(\mathbb{R})$, and $\frac{1}{p}+\frac{1}{q}=1+\frac{1}{s}$. Then,

$$\|f*g\|_s\le\|f\|_p\,\|g\|_q\,, \tag{2.11}$$

In the above, $f*g$ represents the *convolution* of f and g and is the function given by

$$x\mapsto\int_{\mathbb{R}} f(x)g(x-y)\,\mathrm{d}y.$$

In compact spaces, equality holds in Equation (2.11) when f and g are constant functions.
On $\mathbb{R}$ however, we can add a multiplicative constant $c_{p,q}<1$ on the right and improve the inequality. Here, equality holds when f and g are appropriate Gaussians $x\mapsto e^{-ax^2}$ and $x\mapsto e^{-bx^2}$, where a and b are some constants depending on p and q.[16]
Young's inequality is often written in an alternate form. Let r be equal to $1-\frac{1}{s}$. We then have $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=2$. Let h be a function such that $\|h\|_r=1$ and

$$\|(f*g)(h)\|_1=\|f*g\|_s\,\|h\|_r\,.$$

We know that such a h exists by choosing that which satisfies the equality condition in Hölder's inequality.
Therefore, rewriting the above in terms of h,

$$\|(f*g)(h)\|\le\|f\|_p\,\|g\|_q\,\|h\|_r\,.$$

More explicitly,

$$\int\int f(y)g(x-y)h(x)\,\mathrm{d}y\,\mathrm{d}x\le\|f\|_p\,\|g\|_q\,\|h\|_r\,.$$

[16] see this paper by Brascamp and Lieb.

Equivalently,

$$\int\int f(y)g(x-y)h(-x)\,\mathrm{d}y\,\mathrm{d}x \le \|f\|_p \|g\|_q \|h\|_r .$$

Note that $(y)+(x-y)+(-x)=0$. Consider the map from $\mathbb{R}^2 \to \mathbb{R}^3$ given by

$$(x,y) \mapsto (y, x-y, -x).$$

The image of this transformation is equal to

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

Therefore, if $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=2$,

$$\int_H f(u)g(v)h(w) \le \|f\|_p \|g\|_q \|h\|_r .$$

We integrate over a two-dimensional measure on the subspace H.

So this is all well and good, but how is it related to volume ratios? The paper of Brascamp and Lieb mentioned in a footnote previously did more than just say that equality holds when the functions are appropriate Gaussians. It actually *generalized* Young's Convolution Inequality to higher-dimensional spaces and any number of functions. Note that the map from $\mathbb{R}^2$ to $\mathbb{R}^3$ that leads to H is given by

$$x \mapsto (\langle x, v_1\rangle, \langle x, v_2\rangle, \langle x, v_3\rangle),$$

where $v_1=(0,1)$, $v_2=(1,-1)$, and $v_3=(-1,0)$. The generalisation led to the following:

Theorem 2.12. If $(v_i)_1^m$ are vectors in $\mathbb{R}^n$ and $(p_i)_1^m$ are positive numbers satisfying

$$\sum_i \frac{1}{p_i} = n$$

and $(f_i)_1^m$ are non-negative measurable functions on $\mathbb{R}$, then the expression

$$\frac{\displaystyle\int_{\mathbb{R}^n} \prod_{i=1}^m f_i(\langle x, v_i\rangle)}{\displaystyle\prod_{i=1}^m \|f_i\|_{p_i}}$$

is "maximized"[17] when the (f_i) are appropriate Gaussian densities $f_i(x)=e^{-\alpha_i x^2}$, where each α_i depends on the (p_i), (v_i), m, and n.

However, this seems quite unwieldy. The constants α_i are quite difficult to compute since they result from non-linear equations of all the variables. When we talk about convex bodies however, this issue completely disappears and gives a surprising connection back to Fritz John's Theorem!

Theorem 2.13. If $(u_i)_1^m$ are unit vectors in $\mathbb{R}^n$, $(c_i)_1^m$ are positive reals, and $(f_i)_1^m$ are non-negative measurable functions such that

$$\sum_{i=1}^m c_i u_i \otimes u_i = I_n,$$

and

$$\int_{\mathbb{R}^n} \prod_{i=1}^m f_i(\langle x, u_i\rangle)^{c_i} \le \prod_{i=1}^m \left(\int f_i\right)^{c_i}$$

[17] there are degenerate cases for which the maximum is not attained

A couple of things to note here are:

- The maximized value is 1 now! The inequality is sharp.
- The c_i play the role of the $\frac{1}{p_i}$. As observed earlier, the (c_i) sum up to 1 just like the $(\frac{1}{p_i})$ should.
- We replace each f_i with $f_i^{c_i}$ to make it easier to state the equality condition.

When each f_i is equal to $t \mapsto e^{-t^2}$,

$$\begin{aligned}
\int_{\mathbb{R}^n} \prod_{i=1}^m f_i(\langle x, u_i \rangle)^{c_i} &= \int_{\mathbb{R}^n} \exp\left(-\sum_i c_i \langle x, u_i \rangle^2\right) \\
&= \int_{\mathbb{R}^n} \exp(-\|x\|^2) \\
&= \int_{\mathbb{R}^n} \exp\left(-\sum_i x_i^2\right) \\
&= \left(\int e^{-t^2}\right)^n \\
&= \left(\int e^{-t^2}\right)^{\sum_i c_i} \\
&= \prod_{i=1}^m \left(\int f_i\right)^{c_i}
\end{aligned}$$

We now prove Theorem 2.10.

Proof. Let K be a convex body with maximal ellipsoid B_2^n, $(u_i)_1^m$ and $(c_i)_1^m$ be the points and constants as mentioned in Fritz John's Theorem, and C be the polytope defined in Equation (2.10). For each $1 \leq i \leq m$, define $f_i : \mathbb{R} \to \mathbb{R}$ to be the indicator function on $[-1, 1]$. Observe that for $x \in \mathbb{R}^n$, $f_i(\langle x, u_i \rangle)$ is non-zero for every i if and only if $|\langle x, u_i \rangle| \leq 1$ for every i, that is, $x \in C$. Therefore,

$$\begin{aligned}
\mathrm{vol}(C) &= \int_{\mathbb{R}^n} \prod_{i=1}^m f_i(\langle x, u_i \rangle)^{c_i} \\
&\leq \prod_{i=1}^m \left(\int f_i\right)^{c_i} \\
&= \prod_{i=1}^m 2^{c_i} = 2^n,
\end{aligned}$$

which proves our claim. ■

The analogous result of Theorem 2.10 for general convex bodies, as might be expected, says that among convex bodies, the regular solid simplex has the largest volume ratio.