# High-Dimensional Convex Geometry

**Amit Rajaraman**

Last updated January 18, 2021

## Contents

# §1. Introduction

**Definition 1.1.** A subset $S$ of a Euclidean space is said to be *convex* if for any $u_1, \ldots, u_r \in S$ and non-negative $\lambda_1, \ldots, \lambda_r$ 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. We sometimes denote it by $B_\infty^n$ since it is the unit ball under the $\ell_\infty$ norm.
- 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 $(\pm 1, 0, \ldots, 0)$, $(0, \pm 1, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, \pm 1)$. Note that this is the unit ball on the $\ell_1$ 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 $\sqrt{n}$.

More generally, a $k$-simplex is a $k$-dimensional polytope (Definition 1.4) which is the convex hull of its $k+1$ vertices.

**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 $\mathrm{vol}(B_1^n) = 2^n/n!$.

## 1.1. The Euclidean Ball

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

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

### 1.1.1. Finding the Volume

Let us now attempt to calculate $v_n = \mathrm{vol}(B_2^n)$. 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 $\theta$, which is a point on the unit ball that represents the direction of the point. We obviously have $x = r\theta$. The point $\theta$ 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 $\sigma_{n-1}$ 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!

### 1.1.2. Some Surprising Results in Concentration

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* $n$! 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[1] 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[2]. These two points seem to be directly contradictory! However, as might

[1] since we could have equally well taken something other than $x_1$.
[2] 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[3], 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[4] 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 $(\pm 1, \dots, \pm 1)$. Now, construct the ball with center $(\frac{1}{2}, \dots, \frac{1}{2})$ 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,

$$\mathrm{vol}(K) = n v_n \int_{S^{n-1}} \int_0^{r(\theta)} s^{n-1} \, \mathrm{d}s \, \mathrm{d}\sigma = v_n \int_{S^{n-1}} r(\theta)^n \, \mathrm{d}\sigma. \tag{1.3}$$

**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 $\ell_1$ norm). For a general symmetric body $K$, the volume is given by

$$\mathrm{vol}(K) = v_n \int_{S^{n-1}} \|\theta\|_K^{-n} \, \mathrm{d}\sigma_{n-1}(\theta) \tag{1.4}$$

## 1.2. The Cube and other Polytopes

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[5] 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?

### 1.2.1. Banach-Mazur Distance and Spherical Caps

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 $\tilde{L} \subseteq K \subseteq d\tilde{L}$.

[3] The answers to this mathoverflow question might further aid understanding
[4] subject to debate
[5] 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 *$\varepsilon$-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 $\varepsilon$. 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 $\varepsilon$-cap.

**Lemma 1.1** (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 $\sigma_{n-1}$) 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. ∎

**Lemma 1.2** (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 $\sigma_{n-1}$) at most $e^{-n\varepsilon^2/2}$.

*Proof.* Let $\alpha = \cos^{-1}(\varepsilon)$. We may assume that $\alpha \in [0, \pi/2]$. Instead of finding the fraction of area of the spherical cap, we shall instead find the fraction of volume subtended at the center by the cap.
When $\varepsilon \leq \frac{1}{\sqrt{2}}$, note that the entire volume is contained in the ball of radius $\sqrt{1-\varepsilon^2}$ centered at $\varepsilon u$. The fraction of volume of this ball is equal to $\left(\sqrt{1-\varepsilon^2}\right)^n \leq e^{-n\varepsilon^2/2}$.
On the other hand, when $\varepsilon > \frac{1}{\sqrt{2}}$, this entire volume is contained in the ball of radius $\frac{1}{2\varepsilon}$ centered at $\frac{1}{2\varepsilon}u$. The fraction of volume of this ball is equal to $(2\varepsilon)^{-n} \leq e^{-n\varepsilon^2/2}$. ■

### 1.2.2. Bounds on Almost-Spherical Polytopes

**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,

$$K = \{x : |\langle x, v_i\rangle| \leq 1 \text{ for } 1 \leq i \leq m\} \tag{1.5}$$

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 1.5 , 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 $\theta$, there is some $i$ such that
$$\langle \theta, v_i\rangle \geq \frac{1}{d}.$$

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 $\varepsilon$-cap ($\varepsilon = 1/d$ here). Given Lemma 1.2, 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 $4^{-n}$ (by Lemma 1.1), the number of points in a 1-separated set is at most $4^n$.

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 (and thus a corresponding polytope) with at most $4^n$ points.

■

## 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 $n$?

### 1.3.1. The Statement of the Theorem

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 $(\alpha_i)$ 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

$$\sum_i c_i u_i = 0 \tag{1.6}$$

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

$$\sum_i c_i \langle x, u_i \rangle^2 = \|x\|^2 . \tag{1.7}$$

### 1.3.2. Some Consequences of Fritz John's Theorem

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[6]. 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 to get a larger ellipsoid.
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.
Equation (1.7) is equivalent to saying that for all $x \in \mathbb{R}^n$,

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

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

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 (1.7) is written more compactly as

$$\sum_i c_i u_i \otimes u_i = I_n. \tag{1.8}$$

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 $\|u_i\|^2 = 1$[7]. Equating the traces of either side of Equation (1.8), we get

$$\sum_i c_i = n. \tag{1.9}$$

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

[7] 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).

It is easily shown that $\sum_i c_i(u_i - x_0) = 0$ and that each $(u_i - x_0)$ is of unit norm, 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 (1.7) and Equation (1.9), 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,
  $$\begin{aligned} & 0 \leq \sum_i c_i \left(1 - \langle x, u_i \rangle\right) \left(\|x\| + \langle x, u_i \rangle\right) \\ \implies & \sum_i c_i \langle x, u_i \rangle^2 \leq \sum_i c_i \|x\| + (1 - \|x\|) \left\langle x, \sum_i c_i u_i \right\rangle \\ \implies & \sum_i c_i \langle x, u_i \rangle^2 \leq \sum_i c_i \|x\| && \text{(since } \sum_i c_i u_i = 0\text{)} \\ \implies & \|x\| \leq n. && \text{(by Equation (1.7) and Equation (1.9))} \end{aligned}$$

■

Let us now prove Fritz John's Theorem.

### 1.3.3. The Proof

**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 $\partial K$ and positive reals $(c_i)_1^m$ satisfying Equation (1.6) and Equation (1.7). 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 $(\alpha_j)$. 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)[8].
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

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

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 $\alpha_i^2$ 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 $\partial K$ and positive reals $(c_i)_1^m$ satisfying Equation (1.6) and Equation (1.8).

*Proof.* We want to show that there exist unit vectors $(u_i)$ in $\partial K$ 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

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

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

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 this implies $B_2^n$ is not a maximal ellipsoid). Then Lemma 1.7 implies that there exists a matrix $H = (h_{i,j})$ such that the linear map $\varphi$ 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 $u$,

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

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

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

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

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

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 $x \in \partial K \setminus N$, $x^\top Hx \geq \lambda_{\min}\|x\|^2$. That is, for all such $x$,

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

Observe that $\inf_{x\in\partial K\setminus N}\|x\|^2 > 1$.[9] We may also assume that $\lambda_{\min} < 0$, since the claim holds trivially otherwise (we have $\|x\|^2 > 1$). Then, we may set $\delta$ 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 $x \in \partial K \setminus N$,

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

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

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 $(\lambda_i)$ 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

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

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). ∎

[9] 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 $\|x\|^2 \geq 1$). However, this is not possible because $\partial K$ is compact, we have removed a neighbourhood around each contact point $u$, and contact points are the only points in $\partial K$ which have norm 1.

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 from 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 constraint that the Euclidean ball is mapped into $K$. There is a more general theory that (attempts to) solve this problem under different constraints.

# §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 $\theta$, 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 $8R^2B_2^n$ 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 $\phi$ 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 $\theta$ in a spherical cap of radius $t$ about $\phi$, $N(\theta)$ is at most $2t$. Lemma 1.1 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 (1.4). 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

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

It may be shown that if a Hadamard matrix of dimension $n > 2$ exists, then $n$ is a multiple of 4.[10] 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[11] 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

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

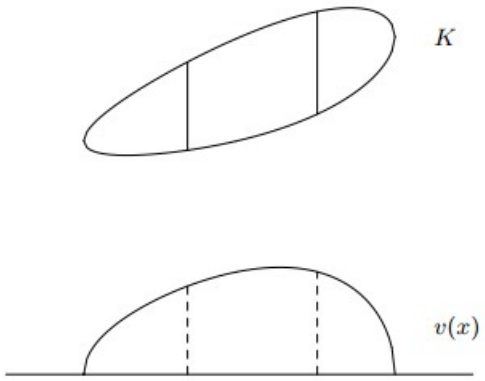
Note that $M$ is a universal constant independent of *everything*. Here the "QS" means "quotient of a subspace".

[10] 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 $n = 4a$.
[11] see Ramsey Theory.

## 2.2. The Prékopa-Leindler Inequality

### 2.2.1. Brunn's Theorem

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 $K$.[12]




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

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

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!

### 2.2.2. The Brunn-Minkowski Inequality

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

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

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

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

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 $B$.

### 2.2.3. The Prékopa-Leindler inequality

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 $\lambda A + (1-\lambda)B$, .[13] 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)!

[13] 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.

**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,

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

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[14]. 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. ■

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

*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,

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

That is, $m_\gamma$, $f_\alpha$, and $g_\beta$ 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 (from [Bus49]) 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

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

Therefore, we shall show that

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

Indeed, we have

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

### 2.3.1. Volume Ratio Estimates and Young's Convolution Inequality

**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 $\operatorname{vol}(K) \leq 2^n$. 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

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

We clearly have $K \subseteq C$, so it suffices to show that $\operatorname{vol}(C) \leq 2^n$.
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,

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

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 $q$.[15]
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 $h$,

[15] see this paper by Brascamp and Lieb.

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

More explicitly,

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

Equivalently,

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

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

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

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

### 2.3.2. A Generalization

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

However, this seems quite unwieldy. The constants $\alpha_i$ 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!

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

**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

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

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.

# §3. Concentration and Almost-Balls

Before we formally begin this section, we state without proof some results in probability that will be helpful later.

Unlike usual convention in probability, we take that a Bernoulli random variable takes $-1$ and $+1$ (instead of 0 and 1) with probability $\frac{1}{2}$ each.

**Theorem 3.1** (Hoeffding's inequality)**.** If $(\varepsilon_i)_1^n$ are independent Bernoulli random variables and $(a_i)_1^n$ are reals that satisfy $\sum_i a_i^2 = 1$, then for $t > 0$,

$$\Pr\left[\left|\sum_{i=1}^n a_i\varepsilon_i\right| > t\right] \leq 2e^{-t^2/2}.$$

We also have

**Theorem 3.2.** If $(X_i)_1^n$ are iid random variables, each of which is uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$ and $(a_i)_1^n$ are reals that satisfy $\sum_i a_i^2 = 1$, then for $t > 0$,

$$\Pr\left[\left|\sum_{i=1}^n a_i X_i\right| > t\right] \leq 2e^{-6t^2}.$$

Given a point $x \in \mathbb{R}^n$, $\sum_i a_i x_i$ is the distance of $x$ from the hyperplane orthogonal to $(a_1, \ldots, a_n) \in \mathbb{R}^n$ that passes through the origin. So the above theorem essentially says that if we uniformly randomly pick a point from $\left[-\frac{1}{2}, \frac{1}{2}\right]^n$, then it is close to any $(n-1)$-dimensional hyperplane passing through the origin. This might be reminiscent of how a majority of the volume in a ball is contained in (relatively) thin slabs.
We elaborate further on this phenomenon in the following section.

## 3.1. Concentration in Geometry

Given a compact set $A \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we write

$$d(x, A) = \inf\{d(x, y) : y \in A\}.$$

Note that for $\varepsilon > 0$,

$$A + \varepsilon B_2^n = \{x \in \mathbb{R}^n : d(x, A) \leq \varepsilon\}.$$

Denote such a neighbourhood $(A + \varepsilon B_2^n)$ of $A$ by $A_\varepsilon$.
Then the proof of the Isoperimetric Inequality we gave using the Brunn-Minkowski inequality essentially says that if $B$ is a Euclidean ball of the same volume as $A$, then

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

Observe that we have removed Minkowski addition and reformulated everything in terms of only the measure and the metric. A more general question that one might ask is:

> Given a metric space $(\Omega, d)$ equipped with a Borel measure $\mu$ and some $\alpha, \varepsilon > 0$, for which sets $A$ of measure $\alpha$ do the "blow-ups" $A_\varepsilon$ have the smallest measure?

### 3.1.1. The Chordal Metric

First, consider the example of $\Omega = S^{n-1}$ and $d$ being the Euclidean metric inherited from $\mathbb{R}^n$ (also known as the *chordal metric*). The measure is $\sigma_{n-1}$.
It was shown (with great difficulty) that the sets $A$ are exactly spherical caps in $\mathbb{R}^n$, which are the balls in $S^{n-1}$.

This might not seem like a big deal, but it does lead to some very startling results. For example, consider some hemisphere $H$ ($\alpha = \frac{1}{2}$). Then for any set $A$ of measure $\frac{1}{2}$, $\sigma(A_\varepsilon) \geq \sigma(H_\varepsilon)$. Further, since the complement of $A$ is a $\varepsilon$-cap, we can use Lemma 1.2 to write

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

This means (for sufficiently large $n$), that nearly the entire sphere lies within distance $\varepsilon$ of $A$, although there might be points that are far from $A$.
Similar to the observation made at the beginning where the majority of the mass was concentrated around any hyperplane through the origin, we see that the majority of the mass is concentrated around any set of measure $\frac{1}{2}$.

Let us now reformulate this same property in another way. Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function:

$$|f(\theta) - f(\phi)| \leq \|\theta - \phi\| \text{ for any } \theta, \phi \in S^{n-1}$$

Let $M \in \mathbb{R}$ (a median of $f$), be such that $\sigma(\{f \geq M\}) = \sigma(\{f \leq M\}) = \frac{1}{2}$. Due to the Lipschitz nature of $f$, for any $\varepsilon > 0$, if $x$ is at distance at most $\varepsilon$ from $\{f \leq M\}$,

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

Writing a similar expression for $\sigma(\{f < M + \varepsilon\})$ and combining the two, we get

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

That is, any 1-Lipschitz function on $S^{n-1}$ is practically constant!

For future reference, we also state two more results. Here, $\operatorname{med}(\cdot)$ represents a *median* of the function, that is, a number $M$ such that $\sigma(\{f \leq M\}) \geq \frac{1}{2}$ and $\sigma(\{f \geq M\}) \leq \frac{1}{2}$.

**Lemma 3.3.** Let $f : S^{n-1} \to \mathbb{R}$ be 1-Lipschitz. Then

$$|\operatorname{med}(f) - \mathbf{E}(f)| \leq 12n^{-1/2}.$$

*Proof.* We have

$$\begin{aligned}
|\operatorname{med}(f) - \mathbf{E}(f)| &\leq \mathbf{E}\left(|f - \operatorname{med}(f)|\right) \\
&\leq \sum_{k=0}^{\infty} \frac{k+1}{\sqrt{n}} \Pr\left[|f - \operatorname{med}(f)| \geq \frac{k}{\sqrt{n}}\right] \\
&\leq n^{-1/2} \sum_{k=0}^{\infty} (k+1) 2e^{-k^2/2} \\
&\leq 12n^{-1/2}.
\end{aligned}$$

■

### 3.1.2. The Gaussian Metric

Second, let us consider the example of $\mathbb{R}^n$ equipped with the standard Gaussian probability measure $\mu$ that has density

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

The solutions to the problem for $\alpha = \frac{1}{2}$ were found to be *half-spaces*. That is, if $A \subseteq \mathbb{R}^n$ and $\mu(A) = \frac{1}{2}$, then for any $\varepsilon > 0$, $\mu(A_\varepsilon) \geq \mu(H_\varepsilon)$, where $H = \{x \in \mathbb{R}^n : x_1 \leq 0\}$ and so, $H_\varepsilon = \{x \in \mathbb{R}^n : x_1 \leq \varepsilon\}$. We have

$$\mu(\overline{H_\varepsilon}) = \frac{1}{\sqrt{2\pi}} \int_\varepsilon^\infty e^{-x^2/2} \, \mathrm{d}x \leq e^{-\varepsilon^2/2}$$

and therefore,

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

A more general result about the Gaussian metric states that

**Theorem 3.4.** Let $A \subseteq \mathbb{R}^n$ be measurable and $\mu$ the standard Gaussian probability measure on $\mathbb{R}$. Then

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

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

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

*Proof.* Define the functions

$$\begin{aligned} f &: x \mapsto e^{d(x,A)^2/4}\gamma(x), \\ g &: x \mapsto \mathbb{1}_A \gamma(x), \text{ and} \\ m &: x \mapsto \gamma(x), \end{aligned}$$

where $\mathbb{1}_A$ represents the indicator function on $A$. Then, for any $x \notin A$ and $y \in A$,

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

Using the above, it is obvious that for any $x, y \in \mathbb{R}^n$,

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

We can then use the Prékopa-Leindler inequality to conclude that

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

That is,

$$\mu(A) \int e^{d(x,A)^2/4} \, \mathrm{d}\mu \leq 1,$$

which is exactly what we want to show.

For the second part, we have

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

For any $\varepsilon > 0$, the integral on the left is at least $e^{\varepsilon^2/4}\mu(\{d(x,A) \geq \varepsilon\})$. We then have

$$\mu(\{d(x,A) \geq \varepsilon\}) \leq 2e^{-\varepsilon^2/4},$$

which directly results in our claim. ∎

The reader might have noticed that we have proved slightly different bounds from what we claimed at the beginning of this subsection ($\varepsilon^2/4$ instead of $\varepsilon^2/2$), we can get arbitrarily close to the given bound by changing $f$, $g$ and $\lambda$ (which we chose to be $\frac{1}{2}$) slightly. Henceforth, we use the $\varepsilon^2/2$ bound itself.

## 3.2. Dvoretzky's Theorem

**Theorem 3.5** (Dvoretzky's Theorem)**.** There is a positive number $c$ such that for every $\varepsilon > 0$ and natural $n$, every symmetric 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 Euclidean ball.

The above theorem essentially says that any symmetric convex body possesses almost spherical slices.
While the original proof was by Dvoretzky, Milman found a different proof of the above theorem which is based on concentration of measure. A few years later, Gordon removed the $\log(1+\varepsilon^{-1})$ factor from the denominator.
We describe Milman's approach without making explicit the dependence on $n$ (for the sake of simplicity).

Loosely, the proof goes as follows:

- Section 3.2.3 - Using Theorem 3.7, restrict to a "good" subspace that is not too much smaller than $\mathbb{R}^n$.
- Section 3.2.4 - Bound the expectation, and thus the median, of the norm corresponding to the body by some constant multiple of $\sqrt{\frac{\log n}{n}}$.
- Section 3.2.1 - Find a general bound on a valid $k$ in terms of the median and use the bound from the previous step.

### 3.2.1. Expressing the Result in Terms of the Median

Let $K$ be a symmetric convex body such that the maximal ellipsoid in it is the Euclidean ball. Let $\|\cdot\|_K$ be the metric under which $K$ is the unit ball.
We then want to find a $k$-dimensional subspace $H$ of $\mathbb{R}^n$ such that the function $f : \theta \mapsto \|\theta\|_K$ is almost constant on the $k$-dimensional ball $H \cap S^{n-1}$. Now, for any $x \in \mathbb{R}^n$, we have $\|x\| \geq \|x\|_K$. This implies that

$$|\|\theta\|_K - \|\phi\|_K| \leq \|\theta - \phi\|_K \leq \|\theta - \phi\|$$

so $f$ is 1-Lipschitz on $S^{n-1}$. From the discussion in Section 3.1.1, we know that on a large part of $S^{n-1}$, $f$ is approximately equal to

$$M = \int_{S^{n-1}} f \, \mathrm{d}\sigma.$$

We can view any such subspace $H$ as an embedding $T : \mathbb{R}^k \to \mathbb{R}^n$. For any unit vector $\psi \in \mathbb{R}^k$, $\|T\psi\|_K$ is close to $M$ with high probability. Then for any unit vectors $(\psi_i)_1^m$, $\|T\psi\|_K$ is close to $M$ with high probability for some choice of $T$. What we would like to show is that if we pin down $\|T\psi_i\|_K$ at sufficiently many points that are "well-distributed" around the ball (in $\mathbb{R}^k$), then the radius will be almost constant on the sphere as well.
To make this more concrete, we bring back some terminology that we used in the proof of Theorem 1.3. Define a set $\{\psi_1, \ldots, \psi_m\}$ to be a *$\delta$-set* in $S^{k-1}$ if for any $x \in S^{k-1}$, $d(x, \psi_i) \leq \delta$ for some $i$.

**Lemma 3.6.** Let $\|\cdot\|_K$ be a norm on $\mathbb{R}^k$. Suppose that for some $\gamma > 0$, each point $\psi$ of some $\delta$-net on $S^{k-1}$ satisfies

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

Then for every $\theta \in S^{k-1}$,

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

*Proof.* We may assume without loss of generality that $M = 1$. Let

$$C = \sup_{\theta \in S^{n-1}} \|\theta\|_K = \sup_{\theta \in S^{n-1}} \frac{\|\theta\|_K}{\|\theta\|}$$

and $\theta_0$ be the point on $S^{n-1}$ for which this is attained. Let $\psi_0$ be a point of the $\delta$-net such that $\|\theta_0 - \psi\| \leq \delta$. Then,

$$\begin{aligned} C &= \|\theta_0\|_K \\ &\leq \|\psi_0 - \theta_0\|_K + \|\psi_0\|_K \\ &\leq C\delta + (1+\gamma). \end{aligned}$$

Therefore,

$$C \leq \frac{1+\gamma}{1-\delta}.$$

Now, for any $\theta \in S^{n-1}$, if $\psi$ is a point of the $\delta$-net such that $\|\theta - \psi\| \leq \delta$, then

$$\begin{aligned} 1 - \gamma &\leq \|\psi\|_K \\ &\leq \|\theta\|_K + \|\psi - \theta\|_K \\ &\leq \|\theta\|_K + \left(\frac{1+\gamma}{1-\delta}\right)\delta. \end{aligned}$$

This directly gives the other side of the inequality. ■

Now, let us bring the problem to the above form.

From the discussion in Section 3.1.1, we know that for any $\gamma > 0$,

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

on all but a set of measure (at most) $2e^{-nM^2\gamma^2/2}$.
We can find a $\delta$-net $\mathcal{A}$ (of the sphere in $\mathbb{R}^{k-1}$ that has at most $\frac{1}{2} \cdot \left(\frac{\delta}{2}\right)^{k-1}$ points.
Now, rather than considering every $k$-dimensional subspace of $\mathbb{R}^n$, we can instead fix a particular embedding of $\mathbb{R}^k$ in $\mathbb{R}^n$ and subsequently consider every orthogonal transformation $U$ of this space.
Rephrasing it in these terms, we want to determine if there is an orthogonal transformation $U$ such that every $\psi \in \mathcal{A}$,

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

For a particular $\psi$, the set of "bad" transformations is of measure at most $2e^{-nM^2\gamma^2/2}$. Further, a necessary (and sufficient) condition for there to be a valid $U$ is that the total set of bad transformations is of measure less than 1. This yields

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

which gives a $k$ in the order of

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

Both the $\gamma$ and the $\delta$ contribute to the $\varepsilon$ as mentioned in Dvoretzky's Theorem, so we should aim to bound $M$ by a reasonably quantity to get the $\log(n)$ estimate we gave in the beginning. In particular, we should show that

$$M = \int_{S^{n-1}} \|\theta\|_K \, \mathrm{d}\sigma \text{ is of the order of } \sqrt{\frac{\log n}{n}}.$$

To show this, we must use the fact that $B_2^n$ is the maximal ellipsoid in $K$. Before we prove the general case, we examine a specific case.

### 3.2.2. Dvoretzky's Theorem for a Cross-Polytope under the Gaussian Measure

Consider the case where the measure is the Gaussian measure on $\mathbb{R}^n$. Then,

$$\begin{aligned} M &= \int_{S^{n-1}} \|\theta\|_K \, \mathrm{d}\sigma \\ &= \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)} \int_{\mathbb{R}^n} \|x\|_K \, \mathrm{d}\mu(x) \qquad \text{(converting from polar coordinates using Equation (1.1))} \\ &> \frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \|x\|_K \, \mathrm{d}\mu(x) \end{aligned}$$

and the body $K$ under consideration is the cross-polytope. The corresponding norm is $x \mapsto \frac{1}{\sqrt{n}} \sum_i |x_i|$. We can split this integral to $n$ separate integrals (in terms of each coordinate) to get

$$M > \frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \|x\|_K \, \mathrm{d}\mu(x) = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, \mathrm{d}x = \sqrt{\frac{2}{\pi}}$$

This proves that for this example, there are in fact almost-spherical sections of dimension of the order of $n$![17]

### 3.2.3. A Weaker Version of the Dvoretzky-Rogers Lemma

To aid us in our goal, we give the following lemma.

**Theorem 3.7** (Dvoretzky-Rogers Lemma)**.** Let $K \subseteq \mathbb{R}^n$ be a symmetric body with maximal ellipsoid $B_2^n$. Then, there exists some subspace $Z \subseteq \mathbb{R}^n$ of dimension $k = \left\lfloor \frac{n}{\log_2 n} \right\rfloor$ and an orthonormal basis $(u_i)_1^k$ of $Z$ such that $\|u_i\|_K \geq \frac{1}{2}$ for all $i$.

The above is quite similar in spirit to Theorem 1.5. Instead of bounding the body between $B_2^n$ and $\sqrt{n}B_2^n$, we bound the restriction of the body to a subspace between a ball and a parallelotope.[18]

The basic idea of the proof is as follows. Either the body $K$ touches the inner ball at a "large number" of places, in which case we can construct the $(u_i)$, or it stays away from the ball in some "large" subspace. In this case, we restrict to this subspace, inflate the inner ball, and repeat the process. Since we cannot inflate the inner ball forever (it must be contained within $\sqrt{n}B_2^n$), this will end at some point.

*Proof.* Consider two cases.

- If every subspace $Y \subseteq \mathbb{R}^n$ of dimension $> n - k$ contains a vector $u$ with $\|u\| = 1$ and $\|u\|_K \geq \frac{1}{2}$, then choose one such vector from each of $k$ orthogonal subspaces of dimension $n - k + 1$ to construct the required.
- Otherwise, there exists some subspace $Y$ of dimension greater than $n-k$ such that for every unit vector $u \in Y$, $\|u\|_K \leq \frac{1}{2}$. Let $\tilde{K}$ and $\tilde{B_2^n}$ be the restriction of $K$ and $B_2^n$ to this subspace respectively. Observe that $2\tilde{B_2^n} \subseteq \tilde{K}$. We can then recurse on the subspace $Y$ scaled down by a factor of $\frac{1}{2}$.

Since $K$ is contained in $\sqrt{n}B_2^n$, this process must terminate in at most $i_0 \coloneqq \log_2(n) - 1$ steps. Because $n - ki_0 \geq k$, the result holds. ∎

Note that we could have repeated the above proof to get a similar result that is off by a constant factor if in Theorem 1.5, we had the weaker bound of something like $n$ or $n^{43}$ instead of $\sqrt{n}$.

While we have found a subspace of dimension $\frac{n}{\log_2 n}$, there are alternate proofs that give much better results. One in particular gives $\dim(Z) = \frac{n}{2}$.

[17] exclamation mark, not factorial!

[18] the parallelotope is defined by the separating hyperplanes between each $(2u_i)$ and $K$.

### 3.2.4. Bounding the Expectation

Henceforth, we restrict ourselves to the subspace $Z$ as defined in Theorem 3.7. We can do so because the reduction in dimension from $n$ to $\frac{n}{\log n}$ makes barely any difference.[19] Assume that $Z = \mathbb{R}^n$.

We now want to get a lower bound on the median of $x \mapsto \|x\|_K$ (restricted to $S^{n-1}$). Instead of doing this, we shall bound the *expectation*. This is justified because as seen in Lemma 3.3, the difference between them is $\mathcal{O}(n^{-1/2})$ which is irrelevant compared to the quantity we aim to bound it by.

Let $\|\cdot\|_P$ be the norm that defines the corresponding parallelotope.
We clearly have $\|\theta\|_K \geq \|\theta\|_P$. Therefore, it suffices to instead bound the expectation of $\|\cdot\|_P$.

Now, we show that it in fact suffices to consider the cube $[-2, 2]^n$ instead of the parallelotope! The norm corresponding to the cube is $\frac{1}{2}\|\cdot\|_\infty$.

**Lemma 3.8.** Let $v_1, \ldots, v_n$ be vectors in a normed space with norm $\|\cdot\|$. Then

$$\sum_{\sigma \in \{-1,1\}^n} \left\| \sum_{i=1}^n \sigma_i v_i \right\| \geq 2^n \max_i \|v_i\|$$

*Proof.* Assume without loss of generality that $v_1$ has the largest norm. Then,

$$\begin{aligned}
\sum_{\sigma \in \{-1,1\}^n} \left\| \sum_{i=1}^n \sigma_i v_i \right\| &= 2 \sum_{\sigma \in \{-1,1\}^{n-1}} \left\| v_1 + \sum_{i=1}^n \sigma_{i-1} v_i \right\| \\
&= 2 \sum_{\sigma \in \{-1,1\}^{n-2}} \left\| v_1 + \left( v_2 + \sum_{i=2}^n \sigma_{i-2} v_i \right) \right\| + \left\| v_1 - \left( v_2 + \sum_{i=2}^n \sigma_{i-2} v_i \right) \right\| \\
&\geq 2 \sum_{\sigma \in \{-1,1\}^{n-2}} 2\|v_1\| \\
&= 2^n \|v_1\| .
\end{aligned}$$

■

For the sake of simplicity, denote the function $x \mapsto \|x\|_P$ by $f_P$ and $x \mapsto \frac{1}{2}\|x\|_\infty$ by $f_C$. We want to show that $\mathbf{E}[f_P] \geq \mathbf{E}[f_C]$. We use an averaging argument as follows, writing each $x \in S^{n-1}$ as $\sum_i \alpha_i u_i$ for some $(\alpha_i)$. We then have

$$\begin{aligned}
2^n \mathbf{E}[f_P] &= 2^n \int_{S^{n-1}} f_P(x)\, \mathrm{d}\sigma(x) \\
&= \sum_{\sigma \in \{-1,1\}^n} \int_{S^{n-1}} f_P \left( \sum_{i=1}^n \sigma_i \alpha_i u_i \right) \mathrm{d}\sigma(x) \\
&= \int_{S^{n-1}} \sum_{\sigma \in \{-1,1\}^n} f_P \left( \sum_{i=1}^n \sigma_i \alpha_i u_i \right) \mathrm{d}\sigma(x) \\
&\geq \int_{S^{n-1}} 2^n \max_i \|\alpha_i u_i\|\, \mathrm{d}\sigma(x) \\
&\geq \int_{S^{n-1}} 2^n \cdot \frac{1}{2} \max_i |\alpha_i|\, \mathrm{d}\sigma(x) \\
&= 2^n \int_{S^{n-1}} f_C(x)\, \mathrm{d}\sigma(x) = 2^n \mathbf{E}[f_C].
\end{aligned}$$

[19] For a more convincing argument, one could just work in the subspace of dimension $\frac{n}{2}$.

Finally, we shall show that for some constant $c$,

$$\mathbf{E}[f_C] = \frac{1}{2}\int_{S^{n-1}} \|x\|_\infty \, \mathrm{d}\sigma(x) \geq c\sqrt{\frac{\log n}{n}}$$

To show this, we give an elegant probabilistic method that uses the fact that the $n$-dimensional Gaussian is symmetric about the origin. Let $Z = (Z_1, \ldots, Z_n)$ be the standard normal. We can draw a point uniformly randomly from $S^{n-1}$ using $\frac{Z}{\|Z\|}$. Clearly,

$$\mathbf{E}[f_C] = \mathbf{E}\left[\frac{\|Z\|_\infty}{\|Z\|}\right].$$

Now, note that

$$\Pr[\|Z\| \geq \sqrt{3n}] = \Pr[\|Z\|^2 \geq 3\mathbf{E}[\|Z\|^2] \leq \frac{1}{3}.$$

That is, $\|Z\|$ is less than some constant multiple of $\sqrt{n}$ with probability at least $\frac{2}{3}$. On the other hand, for any constant $z$,

$$\begin{aligned}
\Pr[\|Z\|_\infty \leq z] &= (\Pr[|Z_1| \leq z])^n \\
&= \left(1 - \int_z^\infty \frac{2}{\sqrt{2\pi}} e^{-t^2/2} \, \mathrm{d}t\right)^n \\
&\leq \left(1 - \frac{2}{\sqrt{2\pi}} e^{-(z+1)^2/2} \, \mathrm{d}t\right)^n
\end{aligned}$$

Setting $z$ to $\sqrt{\log n} - 1$, we get

$$\Pr[\|Z\|_\infty \leq z] \leq \left(1 - \frac{2}{\sqrt{2\pi}} n^{-1/2}\right)^n,$$

which is less than $\frac{1}{3}$ for all $n$. Therefore, $\|Z\|_\infty \leq c_2\sqrt{n}$ with probability at least $\frac{2}{3}$ and $\|Z\| \geq c_1\sqrt{\log n}$ with probability at least $\frac{2}{3}$ for some constants $c_1, c_2$. This implies that $\frac{\|Z\|_\infty}{\|Z\|} \geq c\sqrt{\frac{\log n}{n}}$ for some suitable $c$,[20] which is exactly what we want.

[20] since we know that it occurs with probability at least $\frac{1}{3}$.

# §4. Computing Volume in High Dimensions

A very popular problem in high-dimensional convex geometry is that of determining the volume of an arbitrary body.

For a fixed dimension $n$, this problem isn't too difficult if we want to measure it up to some precision $\varepsilon$. We could assume that the body $K$ is enclosed in a box $B = \bigtimes_{i \le n} [a_i, b_i]$, subdivide this box up to precision $\varepsilon$, and count how many subdivided boxes have non-empty intersection with $K$. This (after normalizing appropriately) can be considered efficient in a fixed dimension, where polynomiality is measure in $\frac{1}{\varepsilon}$. If we measure it in $n$ on the other hand, this method is useless.
We are looking for algorithms that are efficient (polynomial) in the *dimension* $n$.

There are a few issues that arise when we even want to formulate this problem.

- What does it mean when we say that a convex body $K$ is "given"? In what form is it given?
- What does "efficient" exactly mean?

We have already answered the second question above – we are looking for algorithms that are *polynomial* in the dimension $n$. We either want the exact volume or an approximation up to some small *relative error* $\varepsilon > 0$. If it is the latter, we would also like the algorithm to be polynomial in $\frac{1}{\varepsilon}$.[21]

## 4.1. Sandwiching and Deterministic Algorithms

Let us answer the first question that we mentioned – how is an arbitrary body $K$ represented? What information do we have access to?

### 4.1.1. Oracles

We represent the body using an *oracle*. We explain the different types of oracles one may consider over the course of this section.

**Definition 4.1.** A body $K$ is said to be *well-guaranteed* if it contains $rB_2^n$ and is contained in $RB_2^n$ for some $r, R > 0$.

We restrict ourselves to well-guaranteed bodies since otherwise, we may ask any (finite) number of questions about the body (say of the form "is $x \in K$") and receive a `no` every time. This doesn't allow us to make any useful inferences about $\mathrm{vol}(K)$. The fact that $K$ contains $rB_2^n$ ensures that it isn't too small and the containment in $RB_2^n$ ensures that it isn't "at infinity".

A body $K$ is given by an *oracle* $K$ if we know nothing about it other than the fact that it is well-guaranteed with $r$ and $R$ and we may ask questions about $K$, and receive answers to said questions. Depending on the questions and answers, we get different types of oracles.
We primarily use *weak separation oracles* and *strong membership oracles*.

**Definition 4.2** (Strong Membership Oracle)**.** For a fixed convex body $K$, the *strong membership oracle* (correctly) answers questions of the form "Is $x \in K$?".

Now, for $x \notin K$, we know that there is a hyperplane separating them by Lemma 1.7. This gives rise to the strong separation oracle.

**Definition 4.3** (Strong Separation Oracle)**.** For a fixed convex body $K$, the *strong separation oracle* (correctly) answers questions of the form "Is $x \in K$?". If the answer is `no`, it also returns a hyperplane $S$ separating $x$ from $K$.

This hyperplane is returned as a vector $s \in \mathbb{R}^n$ with $\|s\|_\infty = 1$ such that $\langle s, x \rangle > 1$ and $\langle s, y \rangle \le 1$ for any $y \in K$.
This leads to the weak separation oracle.

[21] Note that $\varepsilon$ takes only $\log(1/\varepsilon)$ bits, so this is a relaxation in some sense.

**Definition 4.4** (Weak Separation Oracle)**.** For a fixed convex body $K$, we can fix an $\varepsilon > 0$ and ask the *weak separation oracle* questions of the form "Is $x \in K$ for the positive number $\varepsilon$?". However, in this case, the precision of the answer is $\varepsilon$ in the sense that

(i) If $d(x, \partial K) < \varepsilon$, we can get any answer.

(ii) If $B(x, \varepsilon) \subseteq K$, we get the correct answer (`yes`).

(iii) If $d(x, K) \geq \varepsilon$, we get the correct answer (`no`) and a vector $s$ normalized by $\|s\|_\infty = 1$ for which $\langle s, y\rangle < \langle s, x\rangle + \varepsilon$ for every $y \in K$.

A *weak membership oracle* is similar, where if it is within $\varepsilon$ of $\partial K$, it may return either answer and if it is farther than $\varepsilon$, it returns the correct answer.

The complexity of the algorithm is measured in the number of calls to the oracle, since this is usually the most expensive step.

### 4.1.2. Sandwiching

We earlier mentioned that we only consider well-guaranteed bodies. As might be expected, the ratio $R/r$ is quite important. Defining it slightly more concretely,

**Definition 4.5.** Given a convex body $K$, let $\mathcal{E}$ an ellipsoid centered at 0 such that $\mathcal{E} \subseteq K \subseteq d\mathcal{E}$ for some $d \geq 1$. We are then said to have a *sandwiching* of $K$ with *sandwiching ratio* $d$.

We are given a sandwiching of sandwiching ratio $R/r$ initially. It is natural to want to obtain a sandwiching that has a lower ratio to make whatever algorithm we use more efficient.
Further, note that by Theorem 1.5, the minimum possible sandwiching ratio of (an affine transformation of) a body is at most $n$.

The information given to us initially ($r$ and $R$) are not even necessarily useful all the time. For example, one could have a very "pencil-like" body in $\mathbb{R}^n$ such that the inscribed ball is far far smaller than the circumscribed one. Thus, before we even begin our algorithm, we would want to do some preliminary sandwiching – perform an affine transformation to get a sandwiching with a more manageable sandwiching ratio.

Lovász showed in [Lov86] that it is possible to compute an affine transformation $\tilde{K}$ of $K$ in polynomial time such that

$$B_2^n \subseteq \tilde{K} \subseteq (n+1)\sqrt{n}B_2^n. \tag{4.1}$$

We first introduce the following common tool.

**Lemma 4.1** (Basic ellipsoid method)**.** For a convex body $K \subseteq \mathbb{R}^n$ along with some $R > 0$ such that $K \subseteq RB_2^n$ and a weak separation oracle, it is possible to find a point in $K$ in polynomial time.

We prove it for the case where we have a *strong* separation oracle. The algorithm basically works by cutting down our search space until we find a point.

*Proof.* We construct a sequence $\mathcal{E}_0, \ldots, \mathcal{E}_k$ of ellipsoids with $\mathcal{E}_0 = RB_2^n$. Given $\mathcal{E}_r$, check if its center $x_r$ is contained in $K$. Otherwise, we have a half-space $H_r$ such that $K \subseteq H_r$. We set $\mathcal{E}_{r+1}$ to be the ellipsoid of minimal volume that contains $K \cap H_r$. The sequence terminates when the center of an ellipsoid is contained in $K$.
It may be shown[22] that

$$\mathrm{vol}(\mathcal{E}_{r+1}) = \left(\frac{n}{n+1}\right)^{(n+1)/2}\left(\frac{n}{n-1}\right)^{(n-1)/2}\mathrm{vol}(\mathcal{E}_r).$$

[22] We explicitly give the update formula without proof on the next page. See this.

Rewriting it more suggestively,

$$\begin{aligned}\operatorname{vol}(\mathcal{E}_{r+1}) &= \left(\frac{n^2}{n^2-1}\right)^{(n-1)/2} \frac{n}{n+1} \operatorname{vol}(\mathcal{E}_r) \\ &< \left(1+\frac{1}{n^2}\right)^{(n-1)/2} \operatorname{vol}(\mathcal{E}_r) < e^{-1/2n} \operatorname{vol}(\mathcal{E}_r).\end{aligned}$$

The thing to note here is that $e^{-1/2n}$ is independent of the ellipsoids involved. Since we have $K \subseteq \mathcal{E}_k$,

$$\operatorname{vol}(K) \leq \operatorname{vol}(\mathcal{E}_k) \leq e^{-k/2n}(2R)^n.$$

That is,

$$k \leq 2n^2 \log(2R) - 2n \log(\operatorname{vol}(K))$$

so there is a polynomial upper bound on the number of steps. ■

If each ellipsoid is given by

$$\mathcal{E}_r = \left\{x \in \mathbb{R}^n : (x-x_k)^\top A_k^{-1}(x-x_k) \leq 1\right\},$$

$c_k$ is the vector returned by the separation oracle and $g_k = \frac{1}{\sqrt{c_k^\top A_k c_k}} c_k$, then

$$\begin{aligned}x_{k+1} &= x_k - \frac{1}{n+1} A_k g_k \text{ and} \\ A_{k+1} &= \frac{n^2}{n^2-1}\left(A_k - \frac{2}{n+1} A_k g_k g_k^\top A_k.\right)\end{aligned}$$

Since there is rounding anyway (irrationals might become involved due to the $\sqrt{\cdot}$), it turns out that it suffices to have a weak separation oracle.

A pair of ellipsoids like that in Equation (4.1) is often known as a *weak Löwner-John pair* for $K$ (the sandwiching ratio must be $(n+1)\sqrt{n}$).

**Theorem 4.2.** Let $K \subseteq \mathbb{R}^n$ be a convex body given by a weak separation oracle. Then a weak Löwner-John pair for $K$ can be computed in polynomial time.

Again, we prove it for the case where we have a strong separation oracle instead. This algorithm is nearly identical to that of basic ellipsoid method, but at each step we perform a little extra computation to check if the corresponding ellipsoid scaled down by a factor of $(n+1)\sqrt{n}$ is contained in $K$.

*Proof.* We construct a sequence $\mathcal{E}_0, \ldots, \mathcal{E}_k$ of ellipsoids with $\mathcal{E}_0 = RB_2^n$. Given $\mathcal{E}_r$, first check if its center $x_r$ is contained in $K$. if it is not, then use the basic ellipsoid method to get an ellipsoid that does; we abuse notation and refer to this as $\mathcal{E}_r$ as well.
Next, let the endpoints of the axes of the ellipsoid be given by $x_r \pm a_i$ (for $1 \leq i \leq n$). Check if the $2n$ points $x_r \pm \frac{1}{n+1} a_i$ are in $K$ for each $i$. If they all are, then we are done, since this implies that the convex hull of these points is contained in $K$ as well, and the maximal ellipsoid contained in the convex hull is just $\mathcal{E}_r$ scaled down by a factor of $(n+1)\sqrt{n}$.
Otherwise, suppose that $x_r + \frac{1}{n+1} a_1$ is not in $K$ and $H_r$ is the half-space returned by the oracle that contains $K$. Similar to the basic ellipsoid method, find the minimal ellipsoid $\mathcal{E}_{r+1}$ that contains $\mathcal{E}_r \cap H_r$.
The sequence terminates when we have found a weak Löwner-John pair. ■

One might be tempted to increase the $\frac{1}{n+1}$ factor we use to get something even better, but it is worth noting the reason for both this algorithm working in the first place is that the volume of the ellipsoid decreases at each step.

It is also notable that for certain types of special convex bodies, we can improve the bound beyond $(n+1)\sqrt{n}$. In particular, if $K$ is symmetric, we can attain a factor of $n$, if $K$ is a polytope given as the convex hull of a set of vectors, we can attain $n+1$, and if $K$ is a symmetric polytope given as above, we can attain $\sqrt{n+1}$.

Typically, we assume that after performing sandwiching, we perform a linear transformation such that $B_2^n$ becomes the maximal ellipsoid of the transformed body. That is, the problem boils down computing the volume of a body $K$ with

$$B_2^n \subseteq K \subseteq (n+1)\sqrt{n}B_2^n.$$

### 4.1.3. The Problem and Deterministic Attempts

Our problem is to find for some given convex body $K$, some quantities $\underline{\mathrm{vol}}(K)$ and $\overline{\mathrm{vol}}(K)$ such that

$$\underline{\mathrm{vol}}(K) \leq \mathrm{vol}(K) \leq \overline{\mathrm{vol}}(K)$$

while minimizing $\frac{\overline{\mathrm{vol}}(K)}{\underline{\mathrm{vol}}(K)}$.

Theorem 4.2 produces estimates (equal to the volumes of the ellipsoids) with $\frac{\overline{\mathrm{vol}}(K)}{\underline{\mathrm{vol}}(K)} = n^n(n+1)^{n/2}$. This may seem ludicrously bad, but as it turns out, any deterministic attempts in general are destined to fail. Indeed, Elekes proved in [Ele86] that for any positive $\varepsilon < 2$, there exists no deterministic polynomial time algorithm that returns

$$\frac{\overline{\mathrm{vol}}(K)}{\underline{\mathrm{vol}}(K)} \leq (2-\varepsilon)^n \tag{4.2}$$

for every convex body $K$. The reason for this is that the convex hull of polynomially many points in $B_2^n$ is always bound to be far smaller than $B_2^n$ itself – we've already seen this all the way back in Theorem 1.3. Let us now prove Equation (4.2).

**Lemma 4.3** (Elekes' Theorem)**.** Every deterministic algorithm to estimate the volume of an arbitrary convex body $K \subseteq \mathbb{R}^n$ that uses $q$ oracle queries has $\frac{\overline{\mathrm{vol}}(K)}{\underline{\mathrm{vol}}(K)} \geq \frac{2^n}{q}$ for some $K$ given by a well-guaranteed weak separation oracle.

What exactly do we mean by a deterministic algorithm? Roughly, it means that if we pass the same body into the algorithm twice, we will get the exact same result. More specifically, if we pass two bodies $K_1$ and $K_2$ such that $(x_i)$ and $(y_i)$ are the queried points respectively, the first point where they differ, say $x_i \neq y_i$, must be such that $x_{i-1} = y_{i-1}$ is in $K_1 \triangle K_2$. We abuse this fact.

*Proof.* Let $\mathcal{A}$ be some deterministic algorithm to estimate the volume of a convex body. Fix $\varepsilon = 2^n$ for the separation oracle. When we run $\mathcal{A}$ on $B_2^n$, suppose that the points queried are $x_1, \ldots, x_q$. Let $C$ be the convex hull of these $q$ points. Now, note that if we run $\mathcal{A}$ on $C$, the same points $(x_i)$ will be queried and as a result, the volume estimates $\underline{\mathrm{vol}}$ and $\overline{\mathrm{vol}}$ that are returned are the same as well!
To conclude the argument, note that

$$C \subseteq \bigcup_{i=1}^{q} B\left(\frac{x_i}{2}, \frac{1}{2}\right).$$

Therefore,

$$\frac{\mathrm{vol}(C)}{\mathrm{vol}(B_2^n)} \leq \frac{q}{2^n}.$$

We then have

$$\frac{\overline{\mathrm{vol}}}{\underline{\mathrm{vol}}} \geq \frac{\mathrm{vol}(B_2^n)}{\mathrm{vol}(C)} \geq \frac{2^n}{q}.$$

■

### 4.1.4. The Bárány-Füredi Theorem

We now give a stronger result known as the Bárány-Füredi Theorem (given in [BF87]), which shows that deterministic algorithms in general aren't much better than even the estimate with an $n^n$ error returned by basic sandwiching.

**Theorem 4.4** (Bárány-Füredi Theorem)**.** There is no deterministic polynomial time algorithm that computes a lower bound $\underline{\text{vol}}(K)$ and an upper bound $\overline{\text{vol}}(K)$ for the volume of every convex body $K \subseteq \mathbb{R}^n$ given by some oracle such that for every convex body,

$$\frac{\overline{\text{vol}}(K)}{\underline{\text{vol}}(K)} \leq \left(c\frac{n}{\log n}\right)^n$$

The basic outline of the proof is as follows.

*Proof.* Rather than considering a simple separation oracle, we consider an even stronger oracle. First of all, we know beforehand that the convex body $K$ is such that $B_1^n \subseteq K \subseteq B_\infty^n$.
For $x \in \mathbb{R}^n$, denote $x^\circ = x/\|x\|$, $H^+(x^\circ) = \{z \in \mathbb{R}^d : \langle z, x^\circ\rangle \leq 1\}$ and $H^-(x^\circ) = \{z \in \mathbb{R}^d : \langle z, x^\circ\rangle \geq 1\}$.

When we query $x \in \mathbb{R}^n$, in addition to the information given by the separation oracle, we also receive "$x^\circ \in K$ and $-x^\circ \in K$ and $K \subseteq H^-(x^\circ)$ and $K \subseteq H^+(x^\circ)$". That is, if $x \notin K$, in addition to a separating hyperplane, we also receive information as to whether the hyperplanes at $\pm x^\circ$ that are orthogonal to $x$ are tangential to $K$.

Now, for the main part of the proof, suppose we have some deterministic polyomial time algorithm $\mathcal{A}$ that returns a lower and upper bound $\underline{\text{vol}}(K)$ and $\overline{\text{vol}}(K)$ for any body $K$. The basic idea is roughly similar to that of Elekes' Theorem. Suppose we run $\mathcal{A}$ on $B_2^n$ until $m \leq n^a - n$ questions have been asked for some $a \geq 2$ (due to the polynomial nature of the algorithm) and $x_1, \ldots, x_m$ are the points queried. Define

$$C = \text{Conv}(\pm e_1, \pm e_2, \ldots, \pm e_n, x_1^\circ, \ldots, x_m^\circ).$$

Now, consider the dual $C^*$ of $C$ (recall what a dual is from the proof of Lemma 1.6). Observe that for any of the $x_i$, the output of the oracle on passing $x_i$ (or $\pm e_i$) must be the same whether we pass it with regards to $C$, $C^*$, or $B_2^n$.[23] Indeed, each $H^+(x_i^\circ)$ and $H^-(x_i^\circ)$ is a supporting hyperplane of all three bodies. This implies that the estimates returned by $\mathcal{A}$ are the same for all three bodies!
We then have

$$\overline{\text{vol}}(B_2^n) \geq \text{vol}(C^*) \text{ and } \underline{\text{vol}}(B_2^n) \leq \text{vol}(C).$$

Therefore,

$$\frac{\overline{\text{vol}}(C)}{\underline{\text{vol}}(C)} = \frac{\overline{\text{vol}}(C^*)}{\underline{\text{vol}}(C^*)} = \frac{\overline{\text{vol}}(B_2^n)}{\underline{\text{vol}}(B_2^n)} \geq \frac{\text{vol}(C^*)}{\text{vol}(C)}.$$

■

Over the rest of this section, we show that there is some constant $c$ such that

$$\frac{\text{vol}(C^*)}{\text{vol}(C)} \geq \left(c\frac{n}{\log n}\right)^n.$$

To do this, we introduce some more notation. Let

$$V(n,m) = \sup\{\text{vol}(K) : K = \text{Conv}(\{v_1, \ldots, v_m\}) \subseteq B_2^n\}$$

and

$$S(n,m) = \inf\{\text{vol}(\{x : |\langle x, v_i\rangle| \leq 1 \text{ for each } i\}) : (v_i)_1^m \in \mathbb{R}^n \text{ such that for each } i, \|v_i\| \leq 1\}$$

[23] if $x_i \in B_2^n$, we receive a `yes`. Otherwise, we receive a `no` along with the information that the hyperplanes at the $x_i^\circ$ are tangential.

Clearly, it suffices to show that

$$\frac{S(n,n^a)}{V(n,n^a)} \geq \left(c\frac{n}{\log n}\right)^n \tag{4.3}$$

since $C^*$ and $C$ are of the above considered forms.

### 4.1.5. Bounding $V(n,m)$ and $S(n,m)$

For $1 \leq k \leq n$, define

$$\rho(n,k) = \begin{cases} 1, & \text{if } k = 0 \\ \sqrt{(n-k)/nk}, & \text{if } 1 \leq k \leq n-2 \\ 1/n, & \text{if } k = n-1. \end{cases}$$

**Lemma 4.5.** Let $S = \mathrm{Conv}(\{v_0, v_1, \ldots, v_n\}) \subseteq B_2^n$ be an $n$-dimensional simplex and $x \in S$. Then for every $k$ such that $0 \leq k \leq n-1$, $S$ has a $k$-dimensional face $S_k = \mathrm{Conv}(\{v_{i_0}, v_{i_1}, \ldots, v_{i_k}\})$ and a point $x_k$ in the interior[24] of $S_k$ such that $(x - x_k) \perp \mathrm{span}(S_k)$ and $\|x - x_k\| \leq \rho(n,k)$.

*Proof.* The result for $k = n-1$ follows directly from the fact that the maximal ellipsoid in $S$ is at most $\frac{1}{n}B_2^n$.
For $1 \leq k \leq n-2$, we use strong (backward) induction on $k$.
Let $x_n = x$ and for each $r : n > k > k$, let $x_r$ be such that $(x_{r+1} - x_r) \perp \mathrm{span}(S_r)$ and $\|x_{r+1} - x_r\| \leq \rho(n,r)$.
Note that $x_n - x_{n-1}, x_{n-1} - x_{n-2}, \ldots, x_{k+1} - x_k$ are all orthogonal and $\|x_{r+1} - x_r\| \leq \frac{1}{r}$ for each $r$. We then have

$$\begin{aligned} \|x_n - x_k\|^2 &= \sum_{r=k}^{n-1} \|x_{r+1} - x_r\|^2 \\ &\leq \sum_{r=k+1}^{n} \frac{1}{r^2} \\ &\leq \sum_{r=k+1}^{n} \frac{1}{r(r-1)} \\ &= \frac{1}{k} - \frac{1}{n}. \end{aligned}$$

Finally, the result for $k = 0$ follows from the fact that the $(v_i)$ are contained in $B_2^n$. ∎

Observe that this bound is only tight when $k$ is 1 or $n$. Putting this in a slightly more compact form, let $S \subseteq \mathbb{R}^n$ and $U = \mathrm{span}(S)$. If we define

$$S^\rho = S + \left(U^\perp + \rho B_2^n\right),$$

Lemma 4.5 just says that for some $S_k$, $x \in S_k^{\rho(n,k)}$.
It is also worth noting that if $S$ is convex and $\dim U = k$,

$$\mathrm{vol}(S^\rho) = \mathrm{vol}_k(S) v_{n-k} \rho^{n-k} \tag{4.4}$$

**Theorem 4.6.** There is a constant $c > 0$ such that

$$\frac{V(n,m)}{v_n} \leq \left(c\frac{1 + \log(m/n)}{n}\right)^{n/2}$$

and so,

$$V(n,m) \leq \left(\gamma\frac{\sqrt{1 + \log(m/n)}}{n}\right)^n$$

where $\gamma = \sqrt{2\pi e c}$.

[24] it is a convex combination of the $(v_{i_j})$

If $m/n \to \infty$ and $n/\log(m/n) \to \infty$, then there is a constant $c'$ such that

$$\frac{V(n,m)}{v_n} \leq \left(c' \frac{\log(m/n)}{n}\right)^{n/2}$$

*Proof.* It may be shown that if $K = \text{Conv}(\{v_1, \ldots, v_m\}) \subseteq \mathbb{R}^n$, then $K$ is the union of its $n$-dimensional simplices. That is,

$$K = \bigcup_{i_0 < \cdots < i_n} \text{Conv}(\{v_{i_0}, \ldots, v_{i_n}\}).$$

This allows us to bound the volume of $K$. For any $1 \leq k \leq n-1$, we can write

$$K \subseteq \bigcup_{i_0 < \cdots < i_k} \left\{ S^{\rho(n,k)} : S = \text{Conv}(\{v_{i_0}, \ldots, v_{i_k}\}) \right\}.$$

Bounding the volume,

$$\text{vol}(K) \leq \binom{m}{k+1} \max \left\{ \text{vol}\left(S^{\rho(n,k)}\right) : S = \text{Conv}(\{x_0, \ldots, x_k\}) \subseteq B_2^n \right\}.$$

Using Equation (4.4),

$$\text{vol}(K) \leq \binom{m}{k+1} \cdot v_{n-k} \rho(n,k)^{n-k} \cdot \max \left\{ \text{vol}_k(S) : S = \text{Conv}(\{x_0, \ldots, x_k\}) \subseteq B_2^n \right\}.$$

The right-most quantity is maximum when the body is the $k$-dimensional regular solid simplex, whose volume is $(n+1)^{(n+1)/2}/n^{n/2} n!$. This is easily computed by using induction on dimension and the maximality was briefly mentioned at the end of Section 2.3. So,

$$\begin{aligned}
\text{vol}(K) &\leq \binom{m}{k+1} \cdot \frac{\pi^{(n-k)/2}}{\Gamma\left(\frac{n-k}{2}+1\right)} \left(\frac{n-k}{nk}\right)^{(n-k)/2} \frac{(k+1)^{(k+1)/2}}{k^{k/2} k!} \\
&\leq \binom{m}{k+1} \cdot \left(\frac{2\pi e}{n-k}\right)^{(n-k)/2} \left(\frac{n-k}{nk}\right)^{(n-k)/2} \frac{(k+1)^{(k+1)/2}}{k^{k/2} k!} \\
&= \binom{m}{k+1} \cdot \left(\frac{2\pi e}{nk}\right)^{(n-k)/2} \frac{(k+1)^{(k+1)/2}}{k^{k/2} k!} \\
&\leq \frac{m^{k+1}}{(k+1)!} \cdot \left(\frac{2\pi e}{nk}\right)^{(n-k)/2} \frac{(k+1)^{(k+1)/2}}{k^{k/2} k!} \\
&\leq \left(\frac{em}{k+1}\right)^{k+1} \cdot \left(\frac{2\pi e}{nk}\right)^{(n-k)/2} \left(\frac{e}{k}\right)^k.
\end{aligned}$$

And therefore,

$$\frac{\text{vol}(K)}{v_n} \leq \left(\frac{em}{k+1}\right)^{k+1} n^{k/2} k^{-(n+k)/2}.$$

It remains to choose a suitable value of $k$. For the case when $m/n \to \infty$ and $n/\log(m/n) \to \infty$, we can let $k = \left\lceil \frac{n}{2\log(m/n)} \right\rceil$ to obtain

$$\frac{\text{vol}(K)}{v_n} \leq e^{o(n)} \left(\frac{2e \log(m/n)}{n}\right)^{n/2}.$$

■

Note that the above again leads to an inference similar to that we made in Theorem 1.3 – the volume is comparable only when $m$ is exponentially large.

It remains to bound $S(n,m)$.

To do this, we use the second of the following beautiful results (we state the first for the sake of completeness).

**Theorem 4.7** (Blaschke-Santaló Inequality)**.** Let $K$ be a convex body in $\mathbb{R}^n$ with dual $K^*$. Then

$$\mathrm{vol}(K)\,\mathrm{vol}(K^*) \leq \mathrm{vol}(B_2^n)^2$$

with equality when $K$ is an ellipsoid.

**Theorem 4.8** (Inverse Santaló Inequality)**.** Let $K$ be a convex body in $\mathbb{R}^n$ with dual $K^*$. Then

$$\mathrm{vol}(K)\,\mathrm{vol}(K^*) \geq \frac{4^n}{n!}$$

with equality when $K$ is the regular solid simplex.

For our purposes, it suffices to know that there is some constant $c_2$ such that

$$\mathrm{vol}(K)\,\mathrm{vol}(K^*) \geq \left(\frac{c_2}{n}\right)^n.$$

If we let $K = \{x \in \mathbb{R}^n : |\langle x, v_i\rangle| \leq 1 \text{ for each } i\}$ for some $(v_i)_1^m \in \mathbb{R}^n$ such that $\|v_i\| \leq 1$ for each $i$, then note that $K^* = \mathrm{Conv}(\{v_1, \ldots, v_m\})$.
An upper bound then directly follows from Theorem 4.8 and Theorem 4.6. We get for some constants $c_2$ and $\gamma$,

$$S(n,m) \geq \left(\frac{c_2}{n}\right)^n \left(\frac{n}{\gamma\sqrt{\log(m/n)+1}}\right)^n = \left(\frac{c'}{\sqrt{\log(m/n)+1}}\right)^n \tag{4.5}$$

for some constant $c'$.

Finally, to show the bound mentioned in Equation (4.3), use Theorem 4.6 and Equation (4.5) to get

$$\frac{\overline{\mathrm{vol}}(C)}{\underline{\mathrm{vol}}(C)} \geq \frac{S(n,n^a)}{V(n,n^a)} \geq \left(\frac{c_1}{\gamma a}\frac{n}{\log n}\right)^n$$

which is exactly what we want.

## 4.2. Rapidly Mixing Random Walks

It has now clearly been established beyond doubt that deterministic algorithms will get us nowhere. What if instead, we consider randomized algorithms? That is, we are fine with some small probability, say $\eta$, of getting the incorrect answer? We can do far *far* better in this case.

### 4.2.1. An Issue with High Dimensions and the Solution

Reformulating the problem in this context, we pass some $0 < \eta < 1$, some $\varepsilon > 0$, and a well-guaranteed strong membership oracle[25] of a body $K$, and ask for an estimate $\widetilde{\mathrm{vol}}(K)$ such that with probability at least $1-\eta$,

$$(1-\varepsilon)\widetilde{\mathrm{vol}}(K) \leq \mathrm{vol}(K) \leq (1+\varepsilon)\widetilde{\mathrm{vol}}(K).$$

Henceforth, we assume that the reader has a basic understanding of Markov chains and stationary distributions thereof, at least in the discrete case. In case the reader does not, they can skip ahead to Section 4.4.

A simple method that might come to mind is a Monte Carlo algorithm. Find some box $Q$ in which $K$ is contained, uniformly randomly generate a large number of points in $Q$, and find the fraction of points generated that are in $K$ – this is a good estimate of $\frac{\mathrm{vol}(K)}{\mathrm{vol}(Q)}$.

[25] it is in general not too important which oracle we are given.

However, the issue is one that we emphasised very heavily on in the very first (and quantified to some extent in the previous) section: if we take $K = B_2^n$ and $Q = B_\infty^n$, then $\mathrm{vol}(K)/\mathrm{vol}(Q)$ is extremely (exponentially) small, so it will not work (in polynomial time) at all.

That is, the issue is that $\mathrm{vol}(K)$ is extremely small compared to a box it is contained in. To get around this, there is a surprisingly simple solution.

Rather than considering just $K$, consider some $m+1$ bodies $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K$ (for appropriately large $m$) and for each $i$, estimate $\Lambda_i \coloneqq \mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ (we can then estimate $\mathrm{vol}(K)$ as $\mathrm{vol}(K_0)\prod_i \Lambda_i$).
Usually, we take $K_i = K \cap 2^{i/n} B_2^n$. Note that because $K_i \subseteq 2^{i/n} K_{i-1}$ in this case,

$$\Lambda_i = \frac{\mathrm{vol}(K_i)}{\mathrm{vol}(K_{i-1})} \leq 2$$

is not large at all.
The value of $\mathrm{vol}(K_0)$ is already known. But how do we estimate $\Lambda_i$?
As observed earlier, since $1/\Lambda_i$ is not small, we can just stick with Monte Carlo methods, the basic idea being to somehow generate a uniform random distribution on $K_i$ and find the fraction of points generated within $K_{i-1}$. Here on out, our main interest is just to figure out a way of efficiently uniformly randomly generating points from $K_i$.

To do this, we synthesize a Markov chain whose stationary distribution is the uniform distribution on $K_i$. We run the chain for polynomially many steps, and take the resultant state as a point uniformly randomly generated from $K_i$.

Obviously, we want Markov chains that converge to the stationary distribution very rapidly (in polynomial time) since that is the main part of the algorithm that must be made efficient. To restrict ourselves to finding the uniform distribution *within* $K_i$, we skip any move where the random walk attempts to leave $K_i$.[26] At the same time, we count how often we are in $K_{i-1} \subseteq K_i$.

Succinctly, we use "Multiphase Monte Carlo Markov Chain methods" for volume computation. We call the random walk "rapidly mixing" if it gets sufficiently close to the stationary distribution in polynomially many steps.

Another small change is that we make the random walk *lazy*. That is, if we are at $a \in K$, we stay at $a$ with probability $\frac{1}{2}$, and with probability $\frac{1}{2}$ we choose a random direction $a+v$. If $a+v \in K$, we move there. Otherwise, we stay put. There are two reasons for doing this. The first is that sometimes parity issues arise due to which the stationary distribution might not be the uniform one. The second is that it turns the matrix describing the random walk into a positive semidefinite matrix, which is much easier to analyze.

### 4.2.2. Random Walks on Graphs

Before we move onto the general case, let us define random walks in graphs and study them for a bit. Let $G = (V, E)$ be a connected $d$-regular simple graph with $V = \{1, \ldots, n\}$. A *simple random walk on $G$ with initial state $X_0$* is given by

$$\Pr[X_{t+1} = j \mid X_t = i] = \begin{cases} \frac{1}{2}, & i = j, \\ \frac{1}{2d}, & ij \in E, \\ 0, & \text{otherwise.} \end{cases}$$

It may be shown that irrespective of $X_0$, $\lim_{t\to\infty} \Pr[X_t = i] = \frac{1}{n}$ for any $i$. But to see whether the walk is rapidly mixing, we need to know how fast it converges. Let

$$e_{i,t} = \Pr[X_t = i] - \frac{1}{n}$$

[26] For the hit-and-run strategy we describe later, this is unimportant since it never even tries to leave the body.

be the "excess" probability at time $t$ on vertex $i$. Also, denote $\Pr[X_t = i]$ as $p_i^{(t)}$ for the sake of brevity. Denoting the neighbourhood of $i$ by $\Gamma(i)$,

$$\begin{aligned} e_{i,t+1} &= p_i^{(t+1)} - \frac{1}{n} \\ &= \left(\frac{1}{2}p_i^{(t)} + \frac{1}{2d}\sum_{j\in\Gamma(i)} p_j^{(t)}\right) - \frac{1}{n} \\ &= \frac{1}{2}e_{i,t} + \frac{1}{2d}\sum_{j\in\Gamma(i)} e_{j,t} \\ &= \frac{1}{2d}\sum_{j\in\Gamma(i)} (e_{i,t} + e_{j,t}) \end{aligned} \tag{4.6}$$

To be able to quantify our closeness to the stationary distribution, define

$$d_1(t) = d_1(\tilde{X}, t) = \sum_i |e_{i,t}|$$

and

$$d_2(t) = d_2(\tilde{X}, t) = \sum_i e_{i,t}^2.$$

We call a walk $\tilde{X}$ on $G$ *rapidly mixing* if there exists a polynomial $f$ such that for any $0 < \varepsilon < \frac{1}{3}$ and $t \geq f(\log n)\log(1/\varepsilon)$, $d_1(t) \leq \varepsilon$. However, this doesn't completely make sense right now since if we only have a single graph, $n$ is constant.

**Definition 4.6** (Rapidly Mixing Random Walks)**.** Let $(G_i)_{i\in\mathbb{N}}$ be a sequence of graphs where $G_i$ has $n_i$ vertices and $n_i \to \infty$. We say that the simple random walks on $G_1, G_2, \ldots$ are *randomly mixing* if there is a polynomial $f$ (depending only on the sequence $(G_i)$) such that if $0 < \varepsilon < \frac{1}{3}$ and $t \geq f(\log n_i)\log(1/\varepsilon)$, then $d(\tilde{X}_i, t) \leq \varepsilon$ whenever $\tilde{X}_i$ is a simple random walk on $G_i$.

There are some issues that arise when we want to synthesize a rapidly mixing walk. For example, suppose we have a random walk on $[-1, 1]^n$ and we somehow find ourselves near one of the corners. Then the probability of leaving the corner is extremely low (of the order of $2^{-n}$) at each step, which would greatly hinder the speed of convergence.

### 4.2.3. Conductance and Bounding the Speed of Convergence

In a graph, the analogous event is that we get stuck within some subset of vertices that is highly connected within itself, but not very well-connected to its complement. With this in mind, let us define the conductance of a graph. Let $G = (V, E)$ be a graph and $U \subseteq V$ be non-empty. Then define

$$\Phi_G(U) = \frac{e(U, V \setminus U)}{d|U|}$$

where $e(U, V \setminus U)$ is the number of edges between $U$ and $V$.
$\Phi_G(U)$ gives a measure of the "difficulty" we mentioned earlier. The lower it is, the more difficult it is to leave $U$. We might as well consider only sets $U$ with $|U| \leq \frac{n}{2}$. Thus, we define the *conductance* of $G$ by

$$\Phi_G = \min_{1\leq|U|\leq n/2} \Phi_G(U).$$

In graph theoretic contexts, this quantity is more often known as the *Cheeger's constant* of a graph or its *isoperimetric number*.

For graphs in general, denote $\mathrm{vol}(U) = \sum_{u \in U} d(u)$. Then its conductance is

$$\Phi_G = \min_{\substack{U \subseteq V \\ U \neq \varnothing}} \frac{e(U, V \setminus U)}{\min(\mathrm{vol}(U), \mathrm{vol}(V \setminus U))}.$$

Obviously, $0 \leq \Phi_G \leq 1$ for any graph $G$. The upper bound is only attained when $G$ is the graph containing a single vertex, a single edge, or a triangle. The lower bound is attained only when $G$ is disconnected. If $G$ is large, then the best we can hope for is that $\Phi_G$ is not too much lower than $\frac{1}{2}$.

It is intuitively clear that if a graph has high conductance, then any simple walk will converge quite rapidly. This is stated quantitatively in the following.

**Theorem 4.9.** Every simple random walk on a connected $d$-regular graph $G$ satisfies

$$d_2(t+1) \leq \left(1 - \frac{1}{4}\Phi_G^2\right) d_2(t).$$

In particular,

$$d_2(t) \leq \left(1 - \frac{1}{4}\Phi_G^2\right)^t d_2(0) \leq 2\left(1 - \frac{1}{4}\Phi_G^2\right)^t.$$

We prove this using two other lemmas.

**Lemma 4.10.** For any simple random walk on a connected $d$-regular simple graph on $G$,

$$d_2(t+1) \leq d_2(t) - \frac{1}{2d} \sum_{ij \in E} (e_{i,t} - e_{j,t})^2.$$

*Proof.* Using Equation (4.6) along with the Cauchy-Schwarz inequality,

$$\begin{aligned}
d_2(t+1) &= \frac{1}{4d^2} \sum_{i=1}^n \left( \sum_{j \in \Gamma(i)} e_{i,t} + e_{j,t} \right)^2 \\
&\leq \frac{1}{4d} \sum_{i=1}^n \sum_{j \in \Gamma(i)} (e_{i,t} + e_{j,t})^2 \\
&= \frac{1}{2d} \sum_{ij \in E} (e_{i,t} + e_{j,t})^2 \\
&= \frac{1}{d} \sum_{ij \in E} (e_{i,t}^2 + e_{j,t}^2) - \frac{1}{2d} \sum_{ij \in E} (e_{i,t} - e_{j,t})^2 \\
&= d_2(t) - \frac{1}{2d} \sum_{ij \in E} (e_{i,t} - e_{j,t})^2
\end{aligned}$$

■

**Lemma 4.11.** Suppose weights $x_i$ are assigned to the elements of the vertex set $V = [n]$ satisfying $\sum_i x_i = 0$. Then

$$\sum_{ij \in E} (x_i - x_j)^2 \geq \frac{d}{2} \Phi_G^2 \sum_{i=1}^n x_i^2.$$

Observe that setting $x_i = e_{i,t}$ for each $i$ and substituting the above in Lemma 4.10 directly gives Theorem 4.9.

*Proof.* We may assume without loss of generality that $x_1 \geq x_2 \geq \cdots \geq x_n$. Fix $m = \lceil n/2 \rceil$ and for each $i$, let $y_i = x_i - x_m$. Note that it suffices (and is in fact stronger) to prove the inequality for the $(y_i)$ instead of the $(x_i)$ since

$$\frac{d}{2}\Phi_G^2 \sum_{i=1}^{n}(x_i - x_m)^2 = \frac{d}{2}\Phi_G^2 \sum_{i=1}^{n} x_i^2 + \frac{nd}{2}\Phi_G^2 x_m^2.$$

Also, let

$$u_i = \begin{cases} y_i, & i \leq m, \\ 0, & \text{otherwise}, \end{cases} \quad v_i = \begin{cases} 0, & i \leq m, \\ y_i, & \text{otherwise}. \end{cases}$$

Obviously, it suffices to prove the inequality for the $(u_i)$ and $(v_i)$ since

$$(y_i - y_j)^2 = (u_i - u_j + v_i - v_j)^2 \geq (u_i - u_j)^2 + (v_i - v_j)^2$$

and $\sum_i x_i^2 = \sum_i u_i^2 + \sum_i v_i^2$.
We prove it only for the $(u_i)$. Using the Cauchy-Schwarz inequality,

$$\begin{aligned} 2d\sum_{i=1}^{n} u_i^2 \sum_{ij\in E}(u_i - u_j)^2 &= \sum_{ij\in E} 2(u_i^2 + u_j^2) \sum_{ij\in E}(u_i - u_j)^2 \\ &\geq \sum_{ij\in E}(u_i + u_j)^2 \sum_{ij\in E}(u_i - u_j)^2 \\ &\geq \left(\sum_{ij\in E}(u_i^2 - u_j^2)\right)^2 \end{aligned} \tag{4.7}$$

We aim now to bound the term within the square in the final expression.
Suppose that in every edge $ij \in E$, $i < j$. We can then rewrite the expression as

$$\sum_{ij\in E}(u_i^2 - u_j^2) = \sum_{ij\in E}\sum_{l=i}^{j-1}(u_l^2 - u_{l+1}^2) = \sum_{l=1}^{n}(u_l^2 - u_{l+1}^2)e\left([l], [n] \setminus [l]\right).$$

It is very clear now how the conductance enters the picture. Since we can disregard the terms of the summation after $l = m$, the expression on the right is bounded below by $dl\Phi_G$. That is,

$$\begin{aligned} \sum_{ij\in E}(u_i^2 - u_j^2) &\geq \sum_{l=1}^{m}(u_l^2 - u_{l+1}^2)dl\Phi_G \\ &= d\Phi_G \sum_{l=1}^{m} u_l^2. \end{aligned}$$

Substituting the above in Equation (4.7),

$$\sum_{ij\in E}(u_i - u_j)^2 \geq \frac{d}{2}\Phi_G^2 \sum_{l=1}^{n} u_l^2,$$

which is exactly what we want to show. ∎

Since $d_1(t)^2 \leq nd_2(t)$, we have the following corollary of Theorem 4.9.

**Corollary 4.12.** Every simple random walk on a connected $d$-regular graph $G$ satisfies

$$d_1(t) \leq (2n)^{1/2}\left(1 - \frac{1}{4}\Phi_G^2\right)^{t/2}$$

Note that if $G$ is connected (so $\Phi_G \neq 0$), then for

$$t > 8\Phi_G^{-2}\left(\log(2n) + \log(1/\varepsilon)\right) > \frac{2}{-\log\left(1 - \frac{1}{4}\Phi_G^2\right)}\left(\log(2n) + \log(1/\varepsilon)\right), \tag{4.8}$$

$d_1(t) < \varepsilon$. Thus, we have the following sufficient condition for rapid mixing.

**Lemma 4.13.** Let $(G_i)_{i\in\mathbb{N}}$ be a sequence of regular graphs with $|G_i| = n_i \to \infty$. If there exists $k \in \mathbb{N}$ such that

$$\Phi_{G_i} \geq (\log n_i)^{-k}$$

for sufficiently large $i$, then the simple random walks on $(G_i)$ are rapidly mixing.

The entirety of the discussion thus far has been regarding simple random walks. How would one go about generalizing this to aperiodic reversible random walks on finite sets in general?

**Definition 4.7.** Let $V$ be a finite set and $X$ a random walk on $V$ with transition probabilities $p(u, v)$ such that for each $u$, $p(u, u) \geq \frac{1}{2}$. Let $\lambda$ be the (reversible) stationary distribution that satisfies $\lambda(u)p(u, v) = \lambda(v)p(v, u)$. Also, for $U \subseteq V$, write $\lambda(U) = \sum_{u\in U} \lambda(u)$. The *conductance* of $X$ is then

$$\tilde{\Phi}_X = \min_{\lambda(U)\leq\frac{1}{2}} \frac{\sum_{u\in U}\sum_{v\in V\setminus U} \lambda(u)p(u, v)}{\lambda(U)}.$$

Similar to earlier, the lower the conductance, the higher the probability of getting "stuck" somewhere. Note that the conductance here is half as large as the definition we gave for regular graphs (since in this case, $p(u, v) = 1/2d$ replaces the $1/d$ earlier). That is, if $X$ is a simple random walk on a regular graph,

$$\tilde{\Phi}_X = \frac{1}{2}\Phi_G.$$

As earlier, we can measure the distance from $\lambda$ by

$$d_2(t) = \sum_{v\in V}\left(p_v^{(t)} - \lambda(v)\right)^2.$$

We can then prove the following analogue of Theorem 4.9 (in exactly the same way).

**Theorem 4.14.** Let $X$ be a reversible random walk. Then with the notation above,

$$d_2(t + 1) \leq (1 - \tilde{\Phi}_X^2)d_2(t).$$

In particular,

$$d_2(t) \leq 2(1 - \tilde{\Phi}_X^2)^t.$$

### 4.2.4. An Overview of Random Walks for Uniform Distributions

The basic algorithm used in most algorithms that attempt to solve this problem, which we mentioned at the beginning of this section, was proposed by Dyer, Frieze, and Kannan in [DFK91] and has remained largely unchanged. This algorithm is $\mathcal{O}(n^{23}(\log n)^5\varepsilon^{-2}\log(1/\varepsilon)\log(1/\eta))$. Henceforth, to make things relatively simple, we use the $\mathcal{O}^*$ notation that suppresses any powers of $\log n$ and polynomials of $(1/\varepsilon)$ and $\log(1/\eta)$. With this notation, the algorithm is $\mathcal{O}^*(n^{23})$.

We use a multiphase Monte Carlo algorithm while using random walks to sample. The improvements on this algorithm since its proposal have primarily involved changing the random walk used, using the conductance to bound the mixing time when we are likely to be close to the stationary distribution, and bounding the conductance using isoperimetric inequalities.

There are mainly three different types of random walks used.

**Walking on the Grid.** This is probably the simplest graph. It defines a sufficiently fine grid $\mathbb{L}_\delta$ where each step is of size $\delta$. Suppose we are at $x_t$. At each step, we stay put at $x_t$ with probability $\frac{1}{2}$. Otherwise, we choose a random vector $v$ of the $2n$ possible directions. If $x_t + v \notin K$, we remain at $x_t$ and otherwise, we move to $x_{t+1} = x_t + v$. [DFK91] uses this walk with a value of $\delta$ around $n^{-5/2}$. In [LS90], this was improved to a $\delta$ around $n^{-3/2}$.

**Ball-Steps.** In this random walk, we choose some small step-size $\delta$. We use a lazy random walk but when we try to move, we choose a random $v \in \delta B_2^n$. Similar to the grid, if $x_j + v \notin K$, we remain at $x_j$ and otherwise, we move to $x_{j+1} = x_j + v$. In [KLS97], the value of $\delta$ was around $n^{-1/2}$.

**Hit-and-Run.** Unlike the previous two walks where we had to choose a step-size $\delta$, this walk doesn't need anything of the sort. We choose a random unit vector $v$ from $B_2^n$. We then find the length of the intersection of $\{x+tv : t \in \mathbb{R}\}$ with $K$ and pick a uniformly distributed $x'$ from this segment. It is believed that this walk converges very rapidly.

An issue (in any of the walks) that we must figure out how to rectify is that of getting stuck in some corner (we had given this as motivation for defining the conductance of a random walk).
For example, in the ball-step walk, we can consider the *local conductance*

$$\ell_\delta(x) \coloneqq \frac{\mathrm{vol}(K \cap (x + \delta B_2^n))}{\mathrm{vol}(\delta B_2^n)}$$

and the overall conductance

$$\lambda \coloneqq \frac{1}{\mathrm{vol}(K)} \int_K \ell_\delta(x)\,\mathrm{d}x.$$

In recent times, it has also been a common theme to use *Metropolis chains*, which are defined as follows.

**Metropolis Chain.** Suppose we have a function $f$ on $K$ and a random walk (of any of the above types). We can modify our walk using the same laziness as above, but when we wish to move, we check if $f(x) \geq f(x+v)$ (where $x$ is the original position and $x+v$ is the new proposed position) and

- if `yes`, move to $x+v$.
- if `no`, move to $x+v$ with probability $\frac{f(x)}{f(x+v)}$ (and stay at $x$ otherwise).

If $\int f < \infty$, this produces a random walk with stationary distribution that is proportional to $f(x)$.
So far, we have only tried finding uniform distributions within the body $K$ (and never return a point outside the body $K$). Often, however, we sacrifice this in favour of a distribution that can return a point outside of $K$ with not too high probability (say less than $\frac{1}{2}$) that mixes more rapidly. We detail one such algorithm, similar to that in [DF98], in the following section.

## 4.3. A Modified Grid Walk that Runs in $\mathcal{O}^*(n^8)$

### 4.3.1. A Description of the Walk

The algorithm we describe here uses the "Walking on the Grid" mentioned in the previous section. This involves splitting the body into cubes. To this end, it was observed that if want to sandwich a body between two concentric *cubes* instead of balls, then a ratio of $\mathcal{O}(n)$ can be obtained (instead of the ball-sandwiching ratio of $\mathcal{O}(n^{3/2})$). In particular, [AK91] shows that we can find an affine transformation $\tilde{K}$ of $K$ such that

$$B_\infty^n \subseteq \tilde{K} \subseteq 2(n+1)B_\infty^n.$$

Henceforth, we refer to this $\tilde{K}$ as $K$. So in this case, it is more convenient to consider $K_i = 2^{i/n}$, $0 \le i \le m \coloneqq \lceil n \log_2(2(n+1)) \rceil$ instead of the intersections with the balls we used earlier. That is, at each phase we have two bodies $K$ and $L$ such that

$$K_0 = B_\infty^n \subseteq L \subseteq K \subseteq 2(n+1)B_\infty^n = K_m$$

and

$$L \subseteq K \subseteq 2^{1/n} L.$$

The grid graph over which we design our random walk has vertex set

$$V = \frac{1}{2n}\mathbb{Z}^n \cap K_m.$$

That is, $V$ is the vertex set of the grid graph $P_l^n$ with $l = 8n(n+1)+1$ (having $l^n$ vertices). Denote this graph by $G$ (there is an edge between points whose distance under the $\ell_\infty$ norm is $\frac{1}{2n}$).
We wish to create a rapidly mixing random walk that converges to the stationary distribution on $K$ (or something that could serve the same purpose). Let us now define a distribution on $V$ that is the stationary distribution of a specific random walk.
Consider the function $\varphi_0$ on $\mathbb{R}^n$ defined by

$$\varphi_0(x) = \min\left\{ s \ge 0 : x \in \left(1 + \frac{s}{2n}\right) K \right\}$$

and $\varphi$ defined by $\varphi(x) = \lceil \varphi_0(x) \rceil$. Finally, define $f(x) = 2^{-\varphi(x)}$.
There are a few things to observe that make it apparent why this $f$ is a good choice for our purposes:

- For $x \in K$, $f(x) = 1$.
- If $x, y$ are such that $\|x - y\|_\infty \le \frac{1}{2n}$, then $|\varphi_0(x) - \varphi_0(y)| \le 1$. Indeed,

  $$x = y + (x - y) \in \left(1 + \frac{\varphi_0(y)}{2n}\right) K + \frac{1}{2n} K = \left(1 + \frac{\varphi_0(y) + 1}{2n}\right) K$$

  so $\varphi_0(x) \le \varphi_0(y) + 1$.
- How many $x \in V$ are there such that $\varphi(x) = s > 0$ (so $f(x) = 2^{-s}$)? We must have

  $$x \in \left(1 + \frac{s}{2n}\right) K \setminus \left(1 + \frac{s-1}{2n}\right) K.$$

  The volume of the body on the right is about

  $$\left( \left(1 + \frac{s}{2n}\right)^n - \left(1 + \frac{s-1}{2n}\right)^n \right) \mathrm{vol}(K) < (e^{s/2} - 1)\,\mathrm{vol}(K).$$

  Multiplying by an appropriate factor on either side, the number of points in $V$ in this body is at most $(e^{s/2} - 1) f(K)$. Therefore, for $n > 3$,

  $$f(V) = f(K) + \sum_{s=1}^{\infty} 2^{-s}(e^{s/2} - 1) f(K) < 5 f(K).$$

The above suggests that a Metropolis chain under this function might be exactly what we want – the first point ensures that the resulting distribution is constant on $K$.

What is the Metropolis chain corresponding to $f$ for $G$? It is easily checked that its transition matrix is given by

$$p(x,y) = \begin{cases} \frac{1}{4n}, & xy \in E \text{ and } \varphi(y) \le \varphi(x), \\ \frac{1}{8n}, & xy \in E \text{ and } \varphi(y) = \varphi(x) + 1, \\ 1 - \sum_{z \in \Gamma(x)} p(x,z), & x = y, \\ 0, & \text{otherwise.} \end{cases}$$

It can also easily be checked that this walk is reversible.
Now, let the stationary distribution of this walk be $\lambda$, given by $\lambda(x) = cf(x)$ for a suitable normalizing constant $c$. The third point above ensures that $\lambda(K) > 1/5$ and we don't get points outside of $K$ too often.

There is another issue that we haven't mentioned so far that this walk takes care of. When we have such a walk (in general), we would want to be able to compute the transition probabilities efficiently only at the points where we need it – it would be absurd to store the entire transition matrix all the time. In this example, all we have to do is "carry" the current value of $\varphi$ with us. At most $4n$ appeals to the oracle will give us the values of $\varphi$ at all the neighbours! We can start at a point that we know the value of $\varphi$ of, such as $0 \in B_\infty^n \subseteq K$.

### 4.3.2. Showing Rapid Mixing by Bounding Conductance

The only thing that remains to show now is that it suffices to run the above random walk for a polynomial amount of time to get sufficiently close to the stationary distribution, that is, that the walk is rapidly mixing.
By Theorem 4.14, it suffices to show that this walk has large conductance. To do this, we use the following isoperimetric inequality given in [LS90].

**Theorem 4.15.** Let $M \subseteq \mathbb{R}^n$ and $\mathcal{B}(M)$ be the $\sigma$-field of Borel subsets of $M$. Let $F : \operatorname{Int} M \to \mathbb{R}^+$ be a log-concave function and let $\mu$ be the measure on $\mathcal{B}(M)$ with density $F$

$$\mu(A) = \int_A F$$

for $A \in \mathcal{B}(M)$. Then for $A_1, A_2 \in \mathcal{B}(M)$,

$$\min(\mu(A_1), \mu(A_2)) \leq \frac{\operatorname{diam} M}{d(A_1, A_2)} \mu(M \setminus (A_1 \cup A_2)),$$

where $\operatorname{diam} M = \sup\{\|x - y\| : x, y \in M\}$.

This inequality is slightly loose. The best possible constant on the right is $\frac{1}{2}$ and was proved in [DF98]. We shall show now use this to show that the conductance of our random walk is large.
Let us have $U \subseteq V$ with $0 < \lambda(U) < \frac{1}{2}$ and let $\overline{U} = V \setminus U$. Also, let $\partial U$ be the set of vertices in $\overline{U}$ with at least one neighbour in $U$.
Let $M$ be the union of the cubes of side length $1/2n$ centered at vertices of $V$ ($M$ is a solid cube) and $A_1$ be the union of cubes of side length $1/2n$ centered at vertices of $U$. Let $B$ be the union of cubes of volume $2/(2n)^n$ centered at vertices of $\partial U$ and $A_2 = M \setminus (A_1 \cup B)$. Obviously,

$$\operatorname{diam}(M) = \mathcal{O}(n^{3/2}). \tag{4.9}$$

Then, observe that

$$d(A_1, A_2) \geq \frac{1}{2n} \frac{\sqrt{n}}{2} (2^{1/n} - 1) = \Omega(n^{-3/2}). \tag{4.10}$$

for some suitable positive constant $c_1$. Also, for some positive constant $c_2$,

$$\begin{aligned} \sum_{\substack{u \in U \\ v \in \overline{U}}} \lambda(u) p(u, v) &= \sum_{\substack{u \in U \\ v \in B}} \lambda(v) p(v, u) \\ &\geq \sum_{v \in B} \frac{1}{8n} \lambda(v) = c_2 \frac{\lambda(B)}{n}. \end{aligned}$$

We may assume that $\lambda(B)$ is small. Now, define a measure $\mu$ on $\mathcal{B}(M)$ as in Theorem 4.15 with $F = 2^{-\varphi_1}$, where $\varphi_1$ is the maximal convex function on $M$ bounded above by $\varphi$. Observe that $\lambda(u)$ is always within a constant factor

of the $\mu$-measure of the unit cube centered at $u$. Thus, we have

$$\begin{aligned}\frac{\sum_{u\in U}\sum_{v\in\overline{U}}\lambda(u)p(u,v)}{\lambda(U)} &\geq \frac{c_3}{n}\frac{\mu(B)}{\min\{\lambda(U),\lambda(\overline{U}\setminus\partial U)\}}\\ &= \frac{c_3}{n}\frac{\mu(M\setminus(A_1\cup A_2))}{\min\{\mu(A_1),\mu(A_2)\}}\\ &\geq \frac{c_3}{n}\frac{d(A_1,A_2)}{\operatorname{diam} M}\\ &= \Omega(n^{-4})\end{aligned} \tag{4.11}$$

So what is the total time complexity of the algorithm? Combining Equation (4.11) and Equation (4.8) (or rather, the corresponding result for $d_2(t)$ that does not have the $\log n$ factor), the number of steps in the random walk of each phase of the multiphase Metropolis walk is $\mathcal{O}^*(n^8)$. At each step of the walk, we perform $\mathcal{O}(n)$ oracle queries. Finally, there are $\mathcal{O}^*(n)$ phases. All together, the algorithm is $\mathcal{O}^*(n^{10})$.

In [DF98], a more careful analysis is done to show that this algorithm is in fact $\mathcal{O}^*(n^8)$.[27] More precisely, it is

$$\mathcal{O}\left(n^8\varepsilon^{-2}\log\left(\frac{n}{\varepsilon}\right)\log\left(\frac{1}{\eta}\right)\right).$$

Over the course of the next few sections, we describe a $\mathcal{O}^*(n^7)$ volume estimation algorithm given in [LS93].

## 4.4. Measure-Theoretic Markov Chains and Conductance

### 4.4.1. Some Basic Definitions

**Definition 4.8.** Let $\Omega$ be a non-empty set and $\mathcal{A}$ a $\sigma$-algebra on $\Omega$. For every $u\in\Omega$, let $P_u$ be a probability measure on $\Omega$. Also assume that as a function of $u$, $P_u(A)$ is measurable for any $A\in\mathcal{A}$. We call the triple $(\Omega,\mathcal{A},\{P_u : u\in\Omega\})$ a *Markov scheme*. Together, with an initial distribution $Q_0$ on $\Omega$, this defines a *Markov chain*.

A Markov chain is just a sequence of random variables $w_0, w_1, \ldots$ such that $w_0$ is drawn from $Q_0$ and $w_{i+1}$ is drawn from $P_{w_i}$ (independently of the values of $w_0,\ldots,w_{i-1}$). Therefore,

$$\Pr[w_{i+1}\in A \mid w_1=u_1,\ldots,w_i=u_i] = \Pr[w_{i+1}\in A \mid w_i=u_i] = P_{u_i}(A).$$

Let $f:\Omega\times\Omega\to\mathbb{R}$ be an integrable function (with respect to the product measure $\mu\times\mu$) such that $\int_\Omega f(u,v)\,\mathrm{d}\mu(v)=1$ for all $u\in\Omega$. $f$ then defines a Markov scheme as

$$P_u(A)=\int_A f(u,v)\,\mathrm{d}\mu(v).$$

In this case, $f$ is known as the *transition function* of the Markov scheme. The transition function is said to be *symmetric* if $f(x,y)=f(y,x)$.

A probability measure $Q$ on $\Omega$ is said to be the *stationary distribution* of the Markov scheme if for all $A\in\mathcal{A}$,

$$\int_\Omega P_u(A)\,\mathrm{d}Q(u)=Q(A).$$

This just means that every $w_i$ has the same distribution as that of $Q$.

[27] The conductance is actually $\Omega(n^{-3})$.

Now, consider the inner product space $L^2 = L^2(\Omega, \mathcal{A}, Q)$ with inner product

$$\langle f, g \rangle = \int_\Omega fg \, \mathrm{d}Q.$$

Suppose we have some function $g \in L^2$. Then note that the expectation of $g(w_{i+1})$ (as a function of $w_i = u$) defines a positive linear operator[28] $M : L^2 \to L^2$ by

$$(Mg)(u) = \int_\Omega g(v) \, \mathrm{d}P_u(v).$$

Further note that $(M^k g)(u)$ represents the expectation of $g(w_{i+k})$ given that $w_i = u$.
Now, consider a Markov chain where the first element is drawn from the stationary distribution. Then observe that for any function $g \in L^2$,

$$\begin{aligned}
\mathbf{E}[g(w_i)] &= \mathbf{E}[g(w_0)] = \langle g, 1 \rangle \\
\mathbf{E}[g(w_i)^2] &= \mathbf{E}[g(w_0)^2] = \langle g, g \rangle \\
\mathbf{E}[g(w_i) g(w_{i+k})] &= \mathbf{E}[g(w_0) g(w_k)] = \langle g, M^k g \rangle
\end{aligned}$$

A Markov chain is said to be *time-reversible* if for any $A, B \in \mathcal{A}$, the probability of going from $A$ to $B$ is the same as that of going from $B$ to $A$. That is,

$$\int_B P_u(A) \, \mathrm{d}Q(u) = \int_A P_u(B) \, \mathrm{d}Q(u).$$

It is easy to see that it suffices to have the above for all disjoint sets $A$ and $B$. The above can be rewritten in an even more symmetric fashion as

$$\int_B \int_A 1 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) = \int_A \int_B 1 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u).$$

This is equivalent to saying that for any function $g : \Omega \times \Omega \to \mathbb{R}$ (assuming both sides are well-defined),

$$\int_\Omega \int_\Omega F(u, v) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) = \int_\Omega \int_\Omega F(v, u) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u). \tag{4.12}$$

It is equivalent to say that the operator $M$ is self-adjoint.[29] If the Markov scheme can be described by a transition function $f$ (with respect to $Q$), then time-reversibility is equivalent to the symmetry of $f$.
If the Markov scheme is time-reversible, then for any $g \in L^2$,

$$\begin{aligned}
\langle g, g \rangle - \langle g, Mg \rangle &= \int_\Omega g^2 \, \mathrm{d}Q - \int_\Omega \int_\Omega g(u) g(v) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \\
&= \int_\Omega \int_\Omega g^2(u) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) - \int_\Omega \int_\Omega g(u) g(v) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \\
&= \frac{1}{2} \left( \int_\Omega \int_\Omega (g^2(u) + g^2(v)) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) - \int_\Omega \int_\Omega 2 g(u) g(v) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \right) \quad \text{(by Equation (4.12))} \\
&= \frac{1}{2} \int_\Omega \int_\Omega (g(u) - g(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \geq 0.
\end{aligned} \tag{4.13}$$

Therefore, the spectral radius[30] of $M$ is exactly 1.

**Definition 4.9** (Laziness)**.** A Markov chain is said to be *lazy* if for each $u$,

$$P_u(\{u\}) \geq \frac{1}{2}.$$

[28] a linear operator $A$ such that $\langle Ax, x \rangle \geq 0$ for any $x$.
[29] an operator $A$ such that $\langle Ax, y \rangle = \langle x, Ay \rangle$ for any $x, y$.
[30] the largest absolute value of its eigenvalues.

There are two main, albeit minor and technical, reasons for desiring laziness:

- Sometimes, a lack of laziness can cause parity issues which result in the limit distribution of a chain not converging to the stationary distribution.
- In the time-reversible case, it makes the operator $M$ positive semidefinite, thus making it far easier to analyze.

To see why the latter occurs, note that if $M$ is self-adjoint, then so is $2M - I$ and by a proof exactly like that of Equation (4.13),

$$\langle f, Mf\rangle = \frac{1}{2}\langle f, f\rangle + \frac{1}{2}\langle f, (2M - I)f\rangle \geq 0.$$

Any Markov scheme can be made lazy easily by flipping a (fair) coin at each step and making a move only if it lands on tails.

**Lemma 4.16.** Let $w_1, w_2, \ldots$ be a time-reversible Markov chain generated by a lazy Markov scheme $\mathcal{M}$ with $w_0$ drawn from the stationary distribution $Q$ of $\mathcal{M}$. Then for any function $g \in L^2$,

$$\mathbf{E}[g(w_i)g(w_j)] \geq \mathbf{E}[g(w_i)]\mathbf{E}[g(w_j)] = \mathbf{E}[g(w_0)^2].$$

*Proof.* Assume without loss of generality that $j > i$ and $j - i = k$. Then for any function $h$, the positive semidefiniteness of $M$ implies that

$$\mathbf{E}[h(w_i)h(w_j)] = \langle h, M^k h\rangle \geq 0.$$

Applying this to $(g - \mathbf{E}[g(w_0)])$ yields the result. ■

### 4.4.2. Conductance

**Definition 4.10** (Ergodic Flow)**.** Define the *ergodic flow* $\Phi : \mathcal{A} \to [0, 1]$ of a Markov scheme by

$$\Phi(A) = \int_A P_u(\Omega \setminus A)\,\mathrm{d}Q(u).$$

This just measures how likely $w_1$ is to leave the subset $A$ if $w_0$ is initially drawn from $Q$. Observe that since $Q$ is stationary,

$$\begin{aligned}
\Phi(A) - \Phi(\Omega \setminus A) &= \int_A P_u(\Omega \setminus A)\,\mathrm{d}Q(u) - \int_{\Omega\setminus A} P_u(A)\,\mathrm{d}Q(u) \\
&= Q(A) - \int_A P_u(A)\,\mathrm{d}Q(u) - \int_{\Omega\setminus A} P_u(A)\,\mathrm{d}Q(u) \qquad (\text{since } P_u(\Omega \setminus A) = 1 - P_u(A)) \\
&= Q(A) - \int_\Omega P_u(A)\,\mathrm{d}Q(u) = 0.
\end{aligned}$$

Even conversely, if for some probability distribution $Q'$, the function $\Phi' : \mathcal{A} \to [0, 1]$ defined by

$$A \mapsto \int_A P_u(\Omega \setminus A)\,\mathrm{d}Q(u)$$

is invariant under complementation, then $Q'$ is stationary.

**Definition 4.11.** The *conductance* of the Markov scheme is then defined as

$$\Phi = \inf_{0<Q(A)<1/2} \frac{\Phi(A)}{Q(A)}.$$

For $0 \leq s \leq 1$, the *s-conductance* is defined as

$$\Phi_s = \inf_{s<Q(A)\leq 1/2} \frac{\Phi(A)}{Q(A) - s}.$$

The lower the conductance is, the more likely the Markov chain is to "get stuck" somewhere.

For any $u$, $1 - P_u(\{u\})$ is called the *local conductance* of the Markov chain at $u$. If $Q(u) > 0$,[31] then the local conductance is an upper bound on the conductance.
More generally, let

$$H_t = \{u \in \Omega : P_u(\{u\}) > 1 - t\}$$

and $s = Q(H_t)$. Then

$$\Phi(H_t) = \int_{H_t} P_u(\Omega \setminus H_t)\,\mathrm{d}Q(u) < tQ(H_t).$$

Therefore, the $(s/2)$-conductance is at most $2t$.
The main use of defining conductance is that it is closely related to how fast Markov chains converge to their stationary distribution.
Suppose that $Q_k$ is the distribution in the $k$th step of the chain ($Q_k(A) = \Pr[w_k \in A]$). It turns out that if for all $A \in \mathcal{A}$ such that $Q(A) > 0$, $\Phi(A) > 0$, then $Q_k \to Q$ (in the $\ell_1$ distance[32].). This naturally provides a bound on the speed of convergence.
Let us consider the following particular distance function.

### 4.4.3. A Distance Function

**Definition 4.12.** For $x \in [0, 1]$, consider all measurable functions $g : \Omega \to [0, 1]$ such that

$$\int_\Omega g\,\mathrm{d}Q = x.$$

We then define the *distance function* of $Q$ and $Q_k$ by

$$h_k(x) = \sup_g \int_\Omega g(\mathrm{d}Q_k - \mathrm{d}Q) = \sup_g \int_\Omega g\,\mathrm{d}Q_k - x.$$

For example, it is easily shown that for a finite Markov chain with $N$ states and uniform stationary distribution, $h_k(j/N)$ is the sum of the $j$ largest $\left(Q_k(\omega) - \frac{1}{n}\right)$.

There are a few things to note.

- For any $x$, $0 \le h_k \le 1 - x$. The lower bound is because one can consider the constant function $x$ on $\Omega$. The upper bound is because $\int_\Omega g\,\mathrm{d}Q_k$ is bounded above by 1. In particular, $h_k(1) = 0$.
- $h_k$ is a convex function of $x$. We shall see below in Lemma 4.18 that the supremum in the definition of $h_k$ is attained. Then, for any $a, b, \lambda \in [0, 1]$, set $x = \lambda a + (1 - \lambda)b$ and let $g_1, g_2$ be the functions that attain the supremums for $h_k(a)$ and $h_k(b)$. Then,

$$\sup_g \int_\Omega g\,\mathrm{d}Q_k - x \ge \int_\Omega (\lambda g_1 + (1 - \lambda)g_2)\,\mathrm{d}Q_k = \lambda h_k(a) + (1 - \lambda)h_k(b).$$

This definition might seem quite artificial at the moment, but we hope to give more context to it with the following few lemmas.

**Lemma 4.17.** For every set $A \in \mathcal{A}$ with $Q(A) = x$,

$$-h_k(1 - x) \le Q_k(A) - Q(A) \le h_k(x).$$

[31] it is an atom.
[32] given by $\|f\|_1 = \int_\Omega |f|\,\mathrm{d}Q$

*Proof.* The upper bound is immediate from the definition of the distance function by taking $g = \mathbb{1}_A$ (the indicator function on $A$). The similar upper bound for $\Omega \setminus A$ immediately gives the result. ■

**Lemma 4.18.** For every $0 < x < 1$, there exists a function $G$ that is 0-1 valued except possibly on a $Q$-atom[33] that attains the supremum in the definition of $h_k(x)$.

*Proof.* Let $U \in \mathcal{A}$ such that $Q(U) = 0$ and $Q_k(U)$ is maximum. Let $Q'$ and $Q'_k$ be the restrictions of $Q$ and $Q_k$ to $\Omega \setminus U$. Clearly, the way we have defined $U$ implies that $Q'_k$ is absolutely continuous with respect to $Q'$. Thus, let $\phi$ be the Radon-Nikodym derivative of $Q'_k$ with respect to $Q'$.
Now, let $x \in [0,1]$ and $g : \Omega \to [0,1]$ such that $\int_\Omega g \, \mathrm{d}Q = x$.
For $t \geq 0$, define

$$A_t = U \cup \{u \in \Omega \setminus U : \phi(u) \geq t\} \text{ and } s = \inf\{t \geq 0 : Q(A_t) \leq x\}.$$

Observe that since $A_s = \bigcap_{t<s} A_t$, upper semicontinuity implies that $Q(A_s) \geq x$. Also define

$$A' = \bigcup_{t>s} A_t = U \cup \{u \in \Omega \setminus U : \phi(u) > s\}.$$

Lower semicontinuity implies that $Q(A') \leq x$. We also have that $A' \subseteq A_s$ and for every $u \in A_s \setminus A'$, $\phi(u) = s$.
Now, choose a $B \in \mathcal{A}$ such that $A' \subseteq B \subseteq A_s$, $Q(B) \leq x$, and $Q(B)$ is maximum.
We first show that if $Q(B) = x$, then the indicator function on $B$ suffices. Indeed, in this case,

$$\begin{aligned}
\int_\Omega g \, \mathrm{d}Q_k &= \int_U g \, \mathrm{d}Q_k + \int_{B\setminus U} g\phi \, \mathrm{d}Q + \int_{\Omega\setminus B} g\phi \, \mathrm{d}Q \\
&= \int_U g \, \mathrm{d}Q_k + \int_{B\setminus U} (g-1)\phi \, \mathrm{d}Q + \int_{B\setminus U} \mathrm{d}Q_k + \int_{\Omega\setminus B} g\phi \, \mathrm{d}Q \\
&\leq \int_U \mathrm{d}Q_k + s\int_{B\setminus U} (g-1) \, \mathrm{d}Q + \int_{B\setminus U} \mathrm{d}Q_k + s\int_{\Omega\setminus B} g \, \mathrm{d}Q \qquad (0 \leq g \leq 1 \text{ and } \phi \leq s \ Q\text{-almost everywhere on } \Omega \setminus B) \\
&= Q_k(B) + s\int_{\Omega\setminus U} g \, \mathrm{d}Q - s\int_{B\setminus U} \mathrm{d}Q \\
&= Q_k(B) + s(x - Q(B)) = Q_k(B).
\end{aligned}$$

We also see that the supremum is attained when $g = \mathbb{1}_B$.
Next, assume that $Q(B) < x$. Then for every $W \subseteq A_s \setminus B$, either $Q(W) = 0$ or $Q(W) > x - Q(B)$. That is, the measure on $A' \setminus B$ is concentrated at atoms. Let $V$ be one such atom. As shown above,

$$\int_\Omega g \, \mathrm{d}Q_k \leq Q_k(B) + s(x - Q(B)).$$

To show that this bound is attained, let $g = \mathbb{1}_B + \lambda \mathbb{1}_V$ where $\lambda = (x - Q(B))/Q(V)$. Clearly, $0 \leq g \leq 1$. Further,

$$\int_\Omega g \, \mathrm{d}Q = Q(B) + \lambda Q(V) = x$$

and

$$\int_\Omega g \, \mathrm{d}Q_k = Q_k(B) + \lambda Q_k(V) = Q_k(B) + s(x - Q(B))$$

where the last step follows since $\phi(u) = s$ for all $u \in V \subseteq A_s \setminus A'$. ■

**Lemma 4.19.** If $Q$ is atom-free, then

$$h_k(x) = \sup_{\substack{A\in\mathcal{A} \\ Q(A)=x}} \left(Q_k(A) - Q(A)\right).$$

[33] a $Q$-atom is a set $V \in \mathcal{A}$ such that $Q(V) > 0$ and for any $V' \subseteq V$, either $Q(V') = Q(V)$ or $Q(V') = 0$.

This follows directly from the previous lemma.
Although we did say what a rapidly mixing random walk is earlier in Definition 4.6, we now define it more generally.

**Lemma 4.20.**

First, observe that

$$\sup_x h_k(x) = \sup_{A\in\mathcal{A}} |Q_k(A) - Q(A)| = \frac{1}{2}\|Q_k - Q\|_1 .$$

Let us now get on to the main subject of this section, namely that of bounding the speed of convergence of rapidly mixing Markov chains.

### 4.4.4. Rapidly Mixing Markov Chains

**Definition 4.13** (Rapidly Mixing Markov Chain)**.** A Markov chain is said to be *rapidly mixing* if for some $\theta < 1$, $\sup_x h_k(x)$ is $\mathcal{O}(\theta)^k$.

**Theorem 4.21.** For $k \geq 1$, if $s \leq x \leq 1/2$, then

$$h_k(x) \leq \frac{1}{2}\left(h_{k-1}(x - 2\Phi_s(x-s)) + h_{k-1}(x + 2\Phi_s(x-s))\right)$$

and if $1/2 \leq x \leq 1 - s$, then

$$h_k(x) \leq \frac{1}{2}\left(h_{k-1}(x - 2\Phi_s(1-x-s)) + h_{k-1}(x + 2\Phi_s(1-x-s))\right).$$

*Proof.* We prove the first inequality alone.
By Lemma 4.18, let $A$ be a set such that $Q(A) = x$ and $h_k(x) = Q_k(A) - Q(A)$. Define $g_1, g_2 : \Omega \to [0,1]$ by

$$g_1(u) = \begin{cases} 2P_u(A) - 1, & u \in A, \\ 0, & \text{otherwise}, \end{cases} \qquad g_2(u) = \begin{cases} 1, & u \in A, \\ 2P_u(A), & \text{otherwise}. \end{cases}$$

The functions map into $[0,1]$ because the chain is lazy.
Also, let $x_1 = \int_\Omega g_1 \,\mathrm{d}Q$ and $x_2 = \int_\Omega g_2 \,\mathrm{d}Q$. Observe that $x_1 + x_2 = \int_\Omega 2P_u(A)\,\mathrm{d}Q(u) = 2x$. We have

$$\begin{aligned} h_k(x) &= Q_k(A) - Q(A) \\ &= \frac{1}{2}\left(\left(\int_\Omega g_1 \,\mathrm{d}Q_{k-1} - x_1\right) + \left(\int_\Omega g_2 \,\mathrm{d}Q_{k-1} - x_2\right)\right) \\ &\leq h_{k-1}(x_1) + h_{k-1}(x_2). \end{aligned}$$

We also have

$$x_2 - x = x - x_1 = \int_A (2 - 2P_u(A))\,\mathrm{d}Q(u) = 2\Phi(A) \geq 2\Phi_s(x-s).$$

Then with the above, the concavity of $h_{k-1}$ then implies that

$$h_{k-1}(x_1) + h_{k-1}(x_2) \leq h_{k-1}(x - \Phi_s(x-s)) + h_{k-1}(x + \Phi_s(x-s)),$$

which is what we want. ∎

The next result is analogous to Theorem 4.14 and is our main tool in bounding the speed of convergence using the conductance.

**Theorem 4.22.** Let $0 \le s \le 1/2$ and suppose we have $c_1, c_2$ such that for $s \le x \le 1-s$,

$$h_0(x) \le c_1 + c_2 \min\{\sqrt{x-s}, \sqrt{1-s-x}\}.$$

Then for every $k \ge 0$ and $s \le x \le 1-s$,

$$h_k(x) \le c_1 + c_2 \min\{\sqrt{x-s}, \sqrt{1-s-x}\} \left(1 - \frac{\Phi_s^2}{2}\right)^k.$$

*Proof.* We prove this via induction. It clearly holds for $k = 0$. Suppose that $k \le 1$ and $s \le x \le 1/2$. Using induction,

$$\begin{aligned}
h_k(x) &\le \frac{1}{2}\left(h_{k-1}(x - 2\Phi_s(x-s)) + h_{k-1}(x + 2\Phi_s(x-s))\right) \\
&\le c_1 + \frac{c_2}{2}\left(1 - \frac{\Phi_s^2}{2}\right)^{k-1}\left(\sqrt{x - 2\Phi_s(x-s) - s} + \sqrt{x + 2\Phi_s(x-s) - s}\right) \\
&= c_1 + \frac{c_2}{2}\sqrt{x-s}\left(1 - \frac{\Phi_s^2}{2}\right)^{k-1}\left(\sqrt{1 - 2\Phi_s} + \sqrt{1 + 2\Phi_s}\right) \\
&\le c_1 + \frac{c_2}{2}\sqrt{x-s}\left(1 - \frac{\Phi_s^2}{2}\right)^{k-1}\left(1 - \frac{2\Phi_s}{2} - \frac{4\Phi_s^2}{8} + 1 + \frac{2\Phi_s}{2} - \frac{4\Phi_s^2}{8}\right) \\
&= c_1 + c_2\sqrt{x-s}\left(1 - \frac{\Phi_s^2}{2}\right)^k
\end{aligned}$$

■

Writing the above in a slightly more useful form,

**Corollary 4.23.** Let $M = \sup_A Q_0(A)/Q(A)$. Then for every $A \in \mathcal{A}$,

$$|Q_k(A) - Q(A)| \le \sqrt{M}\left(1 - \frac{\Phi_s^2}{2}\right)^k.$$

*Proof.* Clearly, for any $x$, $h_0(x) \le Mx$. We also have $h_0(x) \le 1 - x$. Therefore,

$$h_0(x) \le \sqrt{Mx(1-x)} \le \sqrt{M}\min\{\sqrt{x}, \sqrt{1-x}\}.$$

Theorem 4.22 then implies the required. ■

### 4.4.5. An Important Inequality involving the operator $M$

A little bit of thought makes it quite clear that to analyze the speed of mixing of Markov chains, the spectrum of the operator $M$ is an important parameter.

**Theorem 4.24.** Let $\mathcal{M}$ be a time-reversible Markov scheme with conductance $\Phi$. Then for every $g \in L^2$ with $\mathbf{E}[g] = 0$,

$$\langle g, Mg\rangle \le \left(1 - \frac{\Phi^2}{2}\right)\|g\|^2.$$

*Proof.* As might be expected, we use Equation (4.13) in this proof. It suffices to show that if $\mathbf{E}[g] = 0$,

$$\int_\Omega \int_\Omega (g(u) - g(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \ge \Phi^2 \|g\|^2.$$

Choose a median $r$ of $Q$, that is, a real number such that $Q(\{x : g(x) > r\}) \le 1/2$ and $Q(\{x : g(x) < r\}) \le 1/2$. Let $h(x) = \max\{g(x) - r, 0\}$. Observe that

$$\int_\Omega \int_\Omega (g(u) - g(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \ge \int_\Omega \int_\Omega (h(u) - h(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u)$$

Therefore, it suffices to bound the quantity on the right suitably. To do this, use the Cauchy-Schwarz inequality to get

$$\int_\Omega \int_\Omega (h(u) - h(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \ge \frac{\left(\int_\Omega \int_\Omega |h^2(u) - h^2(v)| \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u)\right)^2}{\int_\Omega \int_\Omega (h(u) + h(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u)}$$

Now, by definition, $Q(\{h \ge 0\}) \le 1/2$. Therefore,

$$\int_\Omega h^2 \, \mathrm{d}Q \ge \frac{1}{2} \int_\Omega (g(x) - r)^2 = \frac{\|g\|^2 + r^2}{2} \ge \frac{\|g\|^2}{2}.$$

To bound the denominator,

$$\int_\Omega \int_\Omega (h(u) + h(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \le 2 \int_\Omega \int_\Omega (h^2(u) + h^2(v)) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) = 2 \|h\|$$

For each $t$, define $A_t = \{x \in \Omega : h(x)^2 \ge t\}$. Then

$$\begin{aligned}
\int_\Omega \int_\Omega |h^2(u) - h^2(v)| \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) &= 2 \int_\Omega \int_{A(h^2(u))} (h^2(v) - h^2(u)) \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \\
&= 2 \int_\Omega \int_{h^2(u)}^\infty P_u(A(t)) \, \mathrm{d}t \, \mathrm{d}Q(u) && \text{(by Fubini's Theorem)} \\
&= 2 \int_0^\infty \int_{\Omega \setminus A(t)} P_u(A(t)) \, \mathrm{d}Q(u) \, \mathrm{d}t \\
&\ge 2\Phi \int_0^\infty Q(A(t)) \, \mathrm{d}t = 2\Phi \int_\Omega h^2 \, \mathrm{d}Q = 2\Phi \|h\|^2 .
\end{aligned}$$

The result follows directly, since we now have

$$\int_\Omega \int_\Omega (h(u) - h(v))^2 \, \mathrm{d}P_u(v) \, \mathrm{d}Q(u) \ge \frac{4\Phi^2 \|h\|^4}{4 \|h\|^2} = 2\Phi^2 \|h\|^2 \ge \Phi^2 \|g\|^2 ,$$

which is exactly what we set out to show. ■

**Corollary 4.25.** Let $\mathcal{M}$ be a time-reversible Markov scheme with conductance $\Phi$. Then for every $f \in L^2$ with $\mathbf{E}[f] = 0$,

$$\langle f, M^k f \rangle \le \left(1 - \frac{\Phi^2}{2}\right)^k \|f\|^2 .$$

We omit the proof of the above.[34] This quite neatly captures the fact that rapid mixing depends heavily on conductance. Indeed, it implies that $\left\|M^k f\right\| \le (1 - \Phi^2/2)^k \|f\|$, so as time progresses, $f$ "flattens out" and goes closer to 0.

The next inequality can be thought of a central limit theorem style inequality.

**Theorem 4.26.** Let $\mathcal{M}$ be a time-reversible Markov scheme with stationary distribution $Q$ and let $w_1, w_2, \ldots$ be a Markov chain generated by $\mathcal{M}$ with initial distribution $Q$. Let $F \in L^2$ and $\xi = \sum_{i=0}^{T-1} F(w_i)$ for some $T$. Then,

$$\mathbf{Var}[\xi] \le \frac{4T}{\Phi^2} \|F\|^2$$

[34] It may be shown by considering $\tilde{M}$, the restriction of $M$ to the invariant subspace $\mathbf{E}[f] = 0$. By the above lemma, $\left\|\tilde{M}\right\| \le 1 - \Phi^2/2$. Then $\left\|\tilde{M}^k\right\| \le \left\|\tilde{M}\right\|^k$, which imediately gives the result.

*Proof.* We may assume that $\mathbf{E}[\xi] = 0$. Then Theorem 4.24 implies that

$$\begin{aligned}
\mathbf{Var}[\xi] = \mathbf{E}[\xi^2] &= \sum_{0\le i,j\le T-1} \mathbf{E}[F(w_i)F(w_j)] \\
&= T\mathbf{E}[F(w_0)^2] + 2\sum_{0\le i<j\le T-1} \mathbf{E}[F(w_0)F(w_{|i-j|}] \\
&= T\left\|F\right\|^2 + 2\sum_{0\le i<j\le T-1} \mathbf{E}[F(w_0)F(w_{|i-j|}] \\
&= T\left\|F\right\|^2 + 2\sum_{k=1}^{T-1}(T-k)\langle F, M^k F\rangle \\
&\le \left\|F\right\|^2\left(T + 2\sum_{k=1}^{T-1}(T-k)\left(1-\frac{\Phi^2}{2}\right)^k\right) \\
&< 2T\left\|F\right\|^2\sum_{k=0}^{T-1}\left(1-\frac{\Phi^2}{2}\right)^k \le \frac{4T}{\Phi^2}\left\|F\right\|^2.
\end{aligned}$$

■

### 4.4.6. Metropolis Chains

While we have used Metropolis chains previously, let us define them more formally for the sake of completeness.

**Definition 4.14** (Metropolis Chain)**.** Let $\mathcal{M}$ be a time-reversible Markov chain on $(\Omega, \mathcal{A})$ and let $F : \Omega \to \mathbb{R}$ be a non-negative measurable function. Suppose that $\overline{F} = \int_\Omega F\,\mathrm{d}Q$ is finite. Denote by $\mu_F$ the measure with density $F$. We then define the *filtering* of $\mathcal{M}$ by $F$, denoted $\mathcal{M}/F$, as the Markov scheme with transition probabilities

$$P_u^F(A) = \begin{cases} \displaystyle\int_A \min\left\{1, \frac{F(v)}{F(u)}\right\}\mathrm{d}P_u(v), & u \notin A, \\ \displaystyle\int_A \min\left\{1, \frac{F(v)}{F(u)}\right\}\mathrm{d}P_u(v) + \ell(u), & u \in A, \end{cases}$$

where

$$\ell(u) = \int_\Omega \max\left\{0, 1 - \frac{F(v)}{F(u)}\right\}\mathrm{d}P_u(u).$$

And as mentioned, this modified chain converges to a distribution proportional to $F$.

**Theorem 4.27.** If $\mathcal{M}$ is time-reversible, then $\mathcal{M}/F$ is also time-reversible and has stationary distribution $Q_F = (1/\overline{F})\mu_F$.

*Proof.* It suffices to show that for any disjoint measurable sets $A$ and $B$,

$$\int_B P_u^F(A)\,\mathrm{d}Q_F(u) = \int_A P_u^F(B)\,\mathrm{d}Q_F(u),$$

that is,

$$\int_B\int_A \min\left\{1, \frac{F(v)}{F(u)}\right\}\frac{F(u)}{\overline{F}}\,\mathrm{d}P_u(v)\,\mathrm{d}Q(u) = \int_A\int_B \min\left\{1, \frac{F(v)}{F(u)}\right\}\frac{F(u)}{\overline{F}}\,\mathrm{d}P_u(v)\,\mathrm{d}Q(u).$$

Rewriting, we want to show that

$$\int_B\int_A \min\left\{F(u), F(v)\right\}\mathrm{d}P_u(v)\,\mathrm{d}Q(u) = \int_A\int_B \min\left\{F(u), F(v)\right\}\mathrm{d}P_u(v)\,\mathrm{d}Q(u),$$

but this follows from the time-reversibility of $\mathcal{M}$. ■

Before we move on to the main algorithm, we set up some prerequisite results to make the discussion in the subsequent section more natural.

## 4.5. An $\mathcal{O}^*(n^7)$ Algorithm using the Ball-Step

As in nearly all volume estimation algorithms, the basic idea remains the same, the changes being only in the walk. The algorithm described here was originally given in [LS93].
Drawing uniformly randomly from the ball is not too difficult. Letting $\xi_1, \dots, \xi_n$ be iid standard normal distributions and $\eta$ be uniformly distributed in $[0,1]$, we see that

$$v_0 = \left( \eta^{1/n} \frac{\xi_1}{\sqrt{\sum_i \xi_i^2}}, \dots, \eta^{1/n} \frac{\xi_n}{\sqrt{\sum_i \xi_i^2}} \right)$$

is uniformly distributed in $B_2^n$.
On a separate note, instead of the "ball" in ball-step, one could use any other symmetric convex body $G$. An obvious choice is the cube, which is quite convenient to draw points from from a programming perspective. Our analysis shall be done using this general case.

### 4.5.1. The Walk

We modify the ball-walk into a suitable Metropolis chain, with the primary function being quite similar to that we used in the $\mathcal{O}^*(n^8)$ algorithm described in a previous section. Define

$$\phi_K(x) = \min\{t \geq 0 : x \in tK\}$$

and $F_K(x) = e^{-\phi_K(x)}$. Clearly, $0 < F_K \leq 1$. We often refer to these functions as just $\phi$ and $F$ if it is clear what convex body we are talking about. First of all, we can use Equation (1.2) and Equation (1.3) to get

$$\int_{\mathbb{R}^n} F = n \operatorname{vol}(K) \int_0^\infty t^{n-1} e^{-t} \, \mathrm{d}t.$$

That is,

$$\operatorname{vol}(K) = \frac{1}{n!} \int_{\mathbb{R}^n} F. \tag{4.14}$$

Now, define

$$\lambda(s) = \frac{1}{s} \left( \frac{1}{(n-1)!} \int_0^s e^{-t} t^{n-1} \, \mathrm{d}t \right)^{1/n}.$$

**Lemma 4.28.** Let $v$ be randomly distributed in $\mathbb{R}^n$ with density $e^{-\phi(v)}/(n-1)!$. Then

$$H(v) = \lambda(\phi(v))v$$

is uniformly distributed on $K$.

## 4.6. An Isoperimetric Inequality

### 4.6.1. Log-Concave Functions

A function $f : \mathbb{R}^n \to \mathbb{R}^+$ is said to be *log-concave* if for any $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$,

$$f(\lambda x + (1-\lambda)y) \leq f(x)^\lambda f(y)^{1-\lambda}.$$

This just means that $\log f$ is concave. While we did not mention this by name, we discussed similar ideas back in the (multiplicative) Brunn-Minkowski inequality Equation (2.7).

It is quite obvious that if $f$ and $g$ are log-concave functions, then so are $fg$ and $\min\{f, g\}$. The following is far less obvious however.

**Lemma 4.29.** Let $f$ and $g$ be two log-concave functions. If their convolution $h$ defined by $h(x) = \int_{\mathbb{R}^n} g(u)f(x-u)\,\mathrm{d}u$ is well-defined, it is log-concave.

Let $F$ be a non-negative integrable funciton on $\mathbb{R}^n$. As in Theorem 4.15, denote by $\mu_F$ the measure with density $F$. We then get the following corollary.

**Corollary 4.30.** Let $K \subseteq \mathbb{R}^n$ be a convex body and $F : \mathbb{R}^n \to \mathbb{R}$ a log-concave function. Then $\mu_F(x + K)$ is a log-concave function of $x$.

This is quite easily proved by setting $f = F$ and $g = \mathbb{1}_K$ in Lemma 4.29.

Setting $K$ to be a rectangle aligned with the axes having edges of length $\varepsilon$ in $k$ directions and $1/\varepsilon$ in the remaining directions gives

**Corollary 4.31.** Let $F : \mathbb{R}^n \to \mathbb{R}^+$ be a log-concave function with finite integral. Then for any subset $\{x_1, \dots, x_k\}$ of variables, the function

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} F \,\mathrm{d}x_1 \dots \mathrm{d}x_k$$

in the remaining variables is log-concave.

Slightly more generally, setting $f = \mathbb{1}_{K'}$ and $g = \mathbb{1}_K$, we get that the function $x \mapsto \mathrm{vol}((x + K') \cap K)$ is concave.

**Corollary 4.32.** Let $K$ and $K'$ be two convex bodies and $t > 0$. If $\{x \in \mathbb{R}^n : \mathrm{vol}((x + K') \cap K) > t\}$ has an interior point, then it is convex. In particular, for any $0 < s < 1$,

$$K_s = \{x \in K : \mathrm{vol}((x + K) \cap K) \geq s \,\mathrm{vol}(K)\}$$

is a convex body.

### 4.6.2. A Localization Lemma

The main result of this section is the following result, which is an improvement of Theorem 4.15.

**Theorem 4.33.** Let $g$ and $h$ be upper semi-continuous Lebesgue integrable functions on $\mathbb{R}^n$ such that their integrals on $\mathbb{R}^n$ are positive. Then there exist points $a, b \in \mathbb{R}^n$ and a linear function $\ell : [0, 1] \to \mathbb{R}^+$ such that

$$\int_0^1 \ell(t) g((1-t)a + tb)\,\mathrm{d}t > 0 \text{ and } \int_0^1 \ell(t) h((1-t)a + tb)\,\mathrm{d}t > 0$$

Alternatively, this means that if $g$ and $h$ have positive integrals, then there is some truncated cone such that the restriction of $g$ and $h$ to this region have positive integrals.[35] Before we prove this, let us discuss some consequences of the result.

[35] The segment joining $a$ and $b$ is the axis of the truncated cone and $\ell$ represents the radius at the appropriate point.

# References

[AK91] David Applegate and Ravi Kannan. Sampling and integration of near log-concave functions. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, STOC '91, page 156–163, New York, NY, USA, 1991. Association for Computing Machinery.

[BF87] Imre Bárány and Zoltán Füredi. Computing the volume is difficult. *Discrete & Computational Geometry*, 2(4):319–326, Dec 1987.

[Bus49] H. Busemann. A theorem on convex bodies of the Brunn-Minkowski type. *Proceedings of the National Academy of Sciences of the United States of America*, 35(1):27–31, Jan 1949. 16588849[pmid].

[DF98] Martin Dyer and Alan Frieze. Computing the volume of convex bodies: A case where randomness provably helps. *Proc Symp Appl Math*, 44, 02 1998.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, January 1991.

[Ele86] G. Elekes. A geometric inequality and the complexity of computing volume. *Discrete & Computational Geometry*, 1(4):289–292, Dec 1986.

[KLS97] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $\mathcal{O}^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

[Lov86] L. Lovász. Algorithmic theory of numbers, graphs and convexity. In *CBMS-NSF regional conference series in applied mathematics*, 1986.

[LS90] L. Lovász and M. Simonovits. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, SFCS '90, page 346–354 vol. 1, USA, 1990. IEEE Computer Society.

[LS93] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures & Algorithms*, 4(4):359–412, 1993.