Algorithmic Convex Geometry

September 2008

Contents

Chapter 1

Overview

Algorithmic problems in geometry often become tractable with the assumption of convexity. Optimization, volume computation, geometric learning and finding the centroid are all examples of problems which are significantly easier for convex sets.

We will study this phenomenon in depth, pursuing three tracks that are closely connected to each other. The first is the theory of geometric inequalities. We begin with classical topics such as the Brunn-Minkowski inequality, and later deal with more recent developments such as isoperimetric theorems for convex bodies and their extensions to logconcave functions. The second track is motivated by the problem of sampling a geometric distribution by a random walk. Here we will develop some general tools and use them to analyze geometric random walks. The inequalities of the first track play a key role in bounding the rate of convergence of these walks. The last track is the connection between sampling and various algorithmic problems, most notably, that of computing the volume of a convex body (or more generally, integrating a logconcave function). Somewhat surprisingly, random sampling will be a common and essential feature of polynomial-time algorithms for these problems. In some cases, including the volume problem, sampling by a random walk is the *only* known way to get a polynomial-time algorithm.

1.1 Learning by random sampling

We will now see our first example of reducing an algorithmic problem to a random sampling problem. In a typical *learning* problem, we are presented with samples $X^1, X^2, ...$ from the domain of the function, and have to guess

the values $f(X^i)$. After after each guess we are told whether it was right or wrong. The objective is to minimize the number of wrong guesses. One assumes there is an unknown function f belonging to some known restricted class of functions.

As a concrete example, suppose there is a fixed unknown vector $\vec{a} \in \mathbb{R}^n$, and our function f is defined by

$$f(\vec{x}) = \begin{cases} \text{True} & \text{if } \vec{a} \cdot \vec{x} \geq 0 \\ \text{False} & \text{if } \vec{a} \cdot \vec{x} < 0 \end{cases}$$

Assume the right answer has components $a_i \in \{-2^b, ..., 2^b\} \subset \mathbb{Z}$. Consider the following algorithm. At each iteration, choose a random $\vec{a}$ from those that have made no mistakes so far, and use that to make the next guess.

If, on every step, we pick the answer according the majority vote of those $\vec{a}$ which have made no mistake so far, then every mistake would cut down the field of remaining voters by at least a factor of 2. As there are $2^{b(n+1)}$ voters at the outset, you would make at most $(b+1)n$ mistakes.

Exercise 1. *Show that for the randomized algorithm above,*

$$\mathsf{E}(\textit{number of mistakes}) \leq 2(b+1)n.$$

Chapter 2

The Brunn-Minkowski Inequality

In this lecture, we will prove a fundamental geometric inequality – the Brunn-Minkowski inequality. This inequality relates the volumes of convex sets in high-dimensional spaces. Let us first recall the definition of convexity.

Definition 1. *Let $K \subseteq \mathbb{R}^n$. K is a convex set if for any two points $x, y \in K$, and any $0 \leq \lambda \leq 1$, $\lambda x + (1-\lambda)y \in K$.*

To motivate the inequality, consider the following version of cutting a (convex) cake: you pick a point x on the cake, your brother makes a single knife cut and you get the piece that contains x. A natural choice for x is the centroid. For a convex set K, it is

$$x = \frac{1}{\mathsf{Vol}(K)} \int_{y \in K} y \, dy.$$

What is the minimum fraction of the cake that you are guaranteed to get?

For convenience, let K be a convex body whose centroid is the origin. Let $u \in \mathbb{R}^n$ be the normal vector defining the following halfspaces:

$$\begin{aligned} H_1 &= \{v \in \mathbb{R}^n : u \cdot v \geq 0\} \\ H_2 &= \{v \in \mathbb{R}^n : u \cdot v < 0\} \end{aligned}$$

Now we can consider the two portions that u cuts out of K:

$$\begin{aligned} K_1 &= H_1 \cap K \\ K_2 &= H_2 \cap K \end{aligned}$$

We would like to compare $\mathsf{Vol}(K_1)$ and $\mathsf{Vol}(K_2)$ with $\mathsf{Vol}(K)$.

Consider first one dimension – a convex body in one dimension is just a segment on the real number line. It's clear that any cut through the centroid of the segment (i.e. the center) divides the area of the segment into two sides of exactly half the area of the segment.

For two dimensions, this is already a non-trivial problem. Let's consider an isosceles triangle, whose side of unique length is perpendicular to the x axis. If we make a cut through the centroid perpendicular to the x axis, it is readily checked that the volume of the smaller side is $\frac{4}{9}$'ths of the total volume. Is this the least possible in $\mathbb{R}^2$? What about in $\mathbb{R}^n$?

The Brunn-Minkowski inequality will be very useful in answering these questions.

2.1 The inequality

We first define the Minkowski sum of two sets:

Definition 2. *Let $A, B \subseteq \mathbb{R}^n$. The Minkowski sum of A and B is the set*

$$A + B = \{x + y : x \in A, y \in B\}$$

How is the volume of $A + B$ related to the volume of A or B? The Brunn-Minkowski inequality relates these quantities.

Theorem 3 (Brunn-Minkowski)**.** *Let $0 \leq \lambda \leq 1$, and suppose that A, B, and $\lambda A + (1 - \lambda)B$ are measurable subsets of $\mathbb{R}^n$. Then,*

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

Recall that for a measurable set A and a scaling factor λ, we have that:

$$\mathsf{Vol}(\lambda(A)) = \lambda^n \mathsf{Vol}(A).$$

It follows that an equivalent statement of the inequality is the following: for measurable sets A, B and $A + B$ over $\mathbb{R}^n$:

$$\mathsf{Vol}(A + B)^{1/n} \geq \mathsf{Vol}(A)^{1/n} + \mathsf{Vol}(B)^{1/n}.$$

2.1.1 Proving the Brunn-Minkowski inequality

For some intuition, let's first consider the Brunn-Minkowski inequality when A and B are axis-aligned cuboids in $\mathbb{R}^n$. A cuboid in $\mathbb{R}^n$ is a generalization

of the familiar rectangle in two dimensions. An axis-aligned cuboid with side lengths $(a_1, a_2, \ldots a_n)$ is the set

$$A = \{x \in \mathbb{R}^n : l_i \leq x_i \leq l_i + a_i\}$$

for $l = (l_1, \ldots, l_n) \in \mathbb{R}^n$.

Cuboids

Let A be a cuboid with side lengths $(a_1, a_2, \ldots, a_n)$ and B be a cuboid with side lengths $(b_1, b_2, \ldots, b_n)$. Let us prove the Brunn-Minkowski inequality for A and B.

This proof will follow easily because if A and B are cuboids, then $A + B$ is a cuboid with side lengths $(a_1 + b_1, a_2 + b_2, \ldots a_n + b_n)$. Since these are cuboids, it is easy to compute their volumes:

$$\mathsf{Vol}(A) = \prod_{i=1}^{n} a_i, \mathsf{Vol}(B) = \prod_{i=1}^{n} b_i, \mathsf{Vol}(A + B) = \prod_{i=1}^{n} a_i + b_i$$

We want to show that

$$\mathsf{Vol}(A + B)^{1/n} \geq \mathsf{Vol}(A)^{1/n} + \mathsf{Vol}(B)^{1/n}.$$

To this end, consider the ratio between the volumes

$$\begin{aligned}
\frac{\mathsf{Vol}(A)^{1/n} + \mathsf{Vol}(B)^{1/n}}{\mathsf{Vol}(A + B)^{1/n}} &= \frac{\left(\prod_{i=1}^{n} a_i\right)^{1/n} + \left(\prod_{i=1}^{n} b_i\right)^{1/n}}{\left(\prod_{i=1}^{n} a_i + b_i\right)^{1/n}} \\
&= \left(\prod_{i=1}^{n} \frac{a_i}{a_i + b_i}\right)^{1/n} + \left(\prod_{i=1}^{n} \frac{b_i}{a_i + b_i}\right)^{1/n} \\
&\leq \sum_{i=1}^{n} \frac{a_i}{a_i + b_i} + \sum_{i=1}^{n} \frac{b_i}{a_i + b_i} \\
&= 1
\end{aligned}$$

Which proves our claim; the inequality used is the standard inequality between the geometric and the arithmetic mean.

Now that we have the result for cuboids, how can we generalize this to arbitrary measurable sets? The key is that any measurable set can be approximated arbitrarily well with unions of cuboids. We will prove that if $A \cup B$ is the union of a finite number of cuboids, the Brunn-Minkowski inequality holds. We can prove the general result by approximating arbitrary measurable sets with unions of cuboids.

Finite unions of cuboids

We prove the Brunn-Minkowski inequality by induction on the number of disjoint cuboids that $A \cup B$ contain. Note that our result for A and B being cuboids is the base case.

We first translate A so that there exists a cuboid in A fully contained in the halfspace $\{x : x_1 \geq 0\}$ and there exists a cuboid in A fully contained in $\{x : x_1 < 0\}$. Now, translate B so that the following inequality holds:

$$\frac{\mathsf{Vol}(A_+)}{\mathsf{Vol}(A)} = \frac{\mathsf{Vol}(B_+)}{\mathsf{Vol}(B)}$$

where:

$$A_+ = A \cap \{x : x_1 \geq 0\}, A_- = A \setminus A_+$$
$$B_+ = B \cap \{x : x_1 \geq 0\}, B_- = B \setminus B_+$$

Note that we can now proceed by induction on the sets $A_+ + B_+$, $A_- + B_-$, since the number of cuboids in $A_+ \cup B_+$ and $A_- \cup B_-$ is fewer than the number of cuboids in $A \cup B$. This allows us to complete the proof:

$$\begin{aligned}
\mathsf{Vol}(A+B) &\geq \mathsf{Vol}(A_+ + B_+) + \mathsf{Vol}(A_- + B_-) \\
&\geq \left(\mathsf{Vol}(A_+)^{1/n} + \mathsf{Vol}(B_+)^{1/n}\right)^n + \left(\mathsf{Vol}(A_-)^{1/n} + \mathsf{Vol}(B_-)^{1/n}\right)^n \\
&= \mathsf{Vol}(A_+)\left(1 + \left(\frac{\mathsf{Vol}(B)}{\mathsf{Vol}(A)}\right)^{1/n}\right)^n + \mathsf{Vol}(A_-)\left(1 + \frac{\mathsf{Vol}(B)}{\mathsf{Vol}(A)}^{1/n}\right)^n \\
&= \mathsf{Vol}(A)\left(1 + \left(\frac{\mathsf{Vol}(B)}{\mathsf{Vol}(A)}\right)^{1/n}\right)^n
\end{aligned}$$

It follows that

$$\mathsf{Vol}(A+B)^{1/n} \geq \mathsf{Vol}(A)^{1/n} + \mathsf{Vol}(B)^{1/n}$$

which was our desired result.

2.2 Grunbaum's inequality

We are now ready to go back to our original question: what is the smallest volume ratio attainable for the smaller half of a cut through the centroid of any convex body? Somewhat surprisingly, the generalization of the two dimensional case we covered in the introduction is the worst case in high dimensions.

Theorem 4 (Grunbaum's inequality)**.** *Let K be a convex body. Then any half-space defined by a normal vector v that contains the centroid of K contains at least a $\frac{1}{e}$ fraction of the volume of K.*

The overview of the proof is that we will consider the body K and a halfspace defined by v going through the centroid. We will simplify K by performing a symmetrization step – replacing each $(n-1)$-dimensional slice S of K perpendicular to v with an $(n-1)$-dimensional ball of the same volume as S to obtain a new convex body K'. We will verify the convexity of K' by using the Brunn-Minkowski inequality. This does not modify the fraction of volume contained in the halfspace. Next, we will construct a cone from K'. The cone will have the property that the fraction of volume in the halfspace can only *decrease*, so it will still be a lower bound on K. Then we will easily calculate the ratio of the volume of a halfspace defined by a normal vector v cutting through the centroid of a cone, which will give us the desired result.

The two steps of symmetrization and conification can be seen in figure 2.1.

Figure 2.1: The first diagram shows the symmetrization step – each slice of K is replaced with an $(n-1)$ dimensional ball. The second diagram shows how we construct a cone from K'.

Proof. Assume without loss of generality that v is the x_1 axis, and that the centroid of K is the origin. From the convex body K, we construct a new set K'. For all $\alpha \in \mathbb{R}$, we place an $(n-1)$ dimensional ball centered around $c = (\alpha, 0, \ldots 0)$ on the last $(n-1)$ coordinates with radius r_α such that

$$\mathsf{Vol}(B(c, r_\alpha)) = \mathsf{Vol}(K \cap \{x : x_1 = \alpha\}).$$

See figure 2.1 for a visual of the construction. Recall that the volume of a $n-1$ dimensional ball of radius r is $f(n-1)r^{n-1}$ (where f is a function not depending on r and only on n), so that we set

$$r_\alpha = \left(\frac{\mathsf{Vol}(K \cap \{x : x_1 = \alpha\})}{f(n-1)}\right)^{1/(n-1)}$$

We first argue that K' is convex. Consider the function $r : \mathbb{R} \to \mathbb{R}$, defined as $r(\alpha) = r_\alpha$. If r is a concave function (i.e. $\forall x, y \in \mathbb{R}, 0 \leq \lambda \leq 1, r(\lambda x + (1-\lambda)y) \geq \lambda r(x) + (1-\lambda)r(y)$), then K' is convex.

Note that r_α is proportional to $\mathsf{Vol}(K \cap \{x : x_1 = \alpha\})^{1/(n-1)}$. Let $A = K \cap \{x : x_1 = a\}$, and let $B = K \cap \{x : x_1 = b\}$. For some $0 \leq \lambda \leq 1$, let $C = K \cap \{x : x_1 = \lambda a + (1-\lambda)b\}$. r is concave if

$$\mathsf{Vol}(C)^{1/n-1} \geq \lambda \mathsf{Vol}(A)^{1/n-1} + (1-\lambda)\mathsf{Vol}(B)^{1/n-1} \tag{2.1}$$

It is easily checked that

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

so that (2.1) holds by the Brunn-Minkowski inequality, and K' is convex. Let

$$\begin{aligned} K'_+ &= K' \cap \{x : x_1 \geq 0\} \\ K'_- &= K' \cap \{x : x_1 < 0\}. \end{aligned}$$

Then the ratio of volumes on either side of $x_1 = 0$ has not changed, i.e.

$$\begin{aligned} \mathsf{Vol}(K'_+) &= \mathsf{Vol}(K \cap \{x : x_1 \geq 0\}) \\ \mathsf{Vol}(K'_-) &= \mathsf{Vol}(K \cap \{x : x_1 < 0\}). \end{aligned}$$

We argue that

$$\frac{\mathsf{Vol}(K'_+)}{\mathsf{Vol}(K')} \geq \frac{1}{e}$$

and the other side will follow by symmetry. To this end, we transform K' into a cone (see figure 2.1). To be precise, we replace K'_+ with a cone C of equal volume with the same base as K'_+. We replace K'_- with the extension E of the cone C, such that $\mathsf{Vol}(E) = \mathsf{Vol}(K'_-)$. The centroid of K' was at the origin; the centroid of $C \cup E$ can only be pushed to the right along x_1, because r is concave. Let the position of the centroid of $C \cup E$ along x_1 be $\alpha \geq 0$. Then we have that

$$\frac{\mathsf{Vol}(K'_+)}{\mathsf{Vol}(K')} \geq \frac{\mathsf{Vol}((C \cup E) \cap \{x : x_1 \geq \alpha\})}{\mathsf{Vol}(C \cup E)}$$

so a lower bound on the latter ratio implies a lower bound on $\frac{\mathsf{Vol}(K'_+)}{\mathsf{Vol}(K')}$. To this end, we compute the ratio of volume on the right side of a halfspace cutting through the centroid of a cone.

Let the height of the cone $C \cup E$ be h, and the length of the base be R. The centroid of the cone is at position α along x_1 given by:

$$\begin{aligned} \frac{1}{\mathsf{Vol}(C \cup E)} \int_{t=0}^{h} t f(n-1) \left(\frac{tR}{h}\right)^{n-1} dt &= \frac{f(n-1)}{\mathsf{Vol}(C \cup E)} \left(\frac{R}{h}\right)^{n-1} \int_{t=0}^{h} t^n dt \\ &= \frac{f(n-1)}{\mathsf{Vol}(C \cup E)} \cdot \frac{R^{n-1}h^2}{n+1} \end{aligned}$$

where $f(\cdot)$ is the function independent of r such that $f(n)r^n$ gives the volume of an n dimensional ball with radius r. By noting that, for a cone,

$$\mathsf{Vol}(C \cup E) = \frac{f(n-1)R^{n-1}h}{n}$$

we have that $\alpha = \frac{n}{n+1}h$. Now, plugging in to compute the volume of C (the right half of the cone), we have our desired result.

$$\begin{aligned}
\mathsf{Vol}(C) &= \int_{t=0}^{\frac{n}{n+1}h} f(n-1)\left(\frac{tR}{h}\right)^{n-1} dt \\
&= \frac{R^{n-1}f(n-1)}{h^{n-1}} \int_{t=0}^{\frac{n}{n+1}h} t^{n-1} dt \\
&= \left(\frac{n}{n+1}\right)^n \mathsf{Vol}(C \cup E) \\
&\geq \frac{1}{e}\mathsf{Vol}(C \cup E)
\end{aligned}$$

□

Chapter 3

Convex Optimization

3.1 Optimization in Euclidean space

Let $S \subset \mathbb{R}^n$, and let $f : S \to \mathbb{R}$ be a real-valued function. The problem of interest may be stated as

$$\begin{array}{ll} \min & f(x) \\ & x \in S, \end{array} \tag{3.1}$$

that is, find the point $x \in S$ which minimizes by f. For convenience, we denote by x^* a solution for the problem (3.1) and by $f^* = f(x^*)$ the associated cost.

Restricting the set S and the function f to be convex[1], we obtain a class of problems which are polynomial-time solvable in

$$n \quad \text{and} \quad \log\left(\frac{1}{\epsilon}\right),$$

where ϵ defines an optimality criterion such as

$$\|x - x^*\| \leq \epsilon. \tag{3.2}$$

Linear programming (LP) is one important subclass of this family of problems. Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. We state a LP problem as

$$\begin{array}{ll} \min\limits_{x} & c^T x \\ & Ax \geq b. \end{array} \tag{3.3}$$

[1] In fact, one could easily extend to the case of quasi-convex f.

Here, the feasible set is a polyhedron, for many CO problems it is also bounded, i.e., it is a polytope.

Here we illustrate other classical examples.

Example 5. *Minimum weight perfect matching problem; given graph $G = (V, E)$, with costs $\omega : E \to \mathbb{R}$, find*

$$\begin{array}{ll} \min & \sum_{e \in E} c_e x_e \\ & \sum_{e \in \delta(v)} x_e = 1, \quad \textit{for all} \;\; v \in V \\ & x_e \in \{0, 1\}, \quad \textit{for all} \;\; e \in E, \end{array} \tag{3.4}$$

where $\delta(v) = \{e = (i, j) \in E : v = i \;\; or \;\; v = j\}$. Due to the last set of constraints, the so called integrality constraints, (3.4) is not a LP.

The previous problem had integrality constraints which usually increase the difficulty of the problem. One possible approach is to solve a linear relaxation to get (at least) bounds for the original problem. In our particular case,

$$\begin{array}{ll} \min & \sum_{e \in E} c_e x_e \\ & \sum_{e \in \delta(v)} x_e = 1, \quad \textit{for all} \;\; v \in V \\ & 0 \leq x_e \leq 1, \quad \textit{for all} \;\; e \in E. \end{array} \tag{3.5}$$

This problem is clearly an LP, and thus, it can be efficiently solved. Unfortunately, the solution for the LP might not be interesting for the original problem (3.4). For example, it is easy to see that the problem (3.4) is infeasible when $|V| = 3$, while the linear relaxation is not[2].

Surprisingly, it is known that the graph G is bipartite if and only if each vertex of the polytope defined in (3.5) is a feasible solution for (3.4). Thus, for this particular case, the optimal cost coincide for both problems.

For general graphs, we need to add another family of constraints in order to obtain the previous equivalence[3].

$$\sum_{e \in (S, S^c)} x_e \geq 1, \quad \textit{for all} \;\; S \subset V, |S| \;\; \textit{odd}. \tag{3.6}$$

Note that there are an exponential number of constraints in this family. Thus, it is not obvious that it is polynomial-time solvable.

Example 6. *Another convex optimization problem when a polyhedra is intersected with a ball of radius R, i.e., the feasible set is nonlinear.*

[2] Just let $x = (0.5, 0.5, 0.5)$.

[3] The proof is due to Jack Edmonds.

$$\begin{array}{ll} \min & c^T x \\ & Ax \geq b \\ & \sum_{i=1}^n x_i^2 \leq R^2. \end{array} \tag{3.7}$$

Example 7. *Semi-definite programming (SDP). Here the variable, instead of being a vector in $\mathbb{R}^n$, is constrained to being a semi-definite positive matrix.*

$$\begin{array}{ll} \min & C \cdot X \\ & A \cdot X \leq b_i, \quad \textit{for } \ i = 1, \ldots, m \\ & X \succcurlyeq 0, \end{array} \tag{3.8}$$

where $A \cdot B = \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij}$ denotes a generalized inner product, and the last constraints represent

$$\forall y \in \mathbb{R}^n, y^T X y \geq 0$$

which defines an infinite number of linear constraints.

One might notice that it is impossible to explicitly store all constraints in (3.6) or (7). It turns out that there is an extremely convenient way to describe the input of a convex programming problem in an implicit way.

Definition 8. *A separation oracle for a convex set $K \subset \mathbb{R}^n$ is an algorithm that for each $x \in \mathbb{R}^n$, either states that $x \in K$, or provides a vector $a \in \mathbb{R}^n$ such that*

$$a^T x < b \ \ \textit{and} \ \ a^T y \geq b \quad \textit{for all} \ \ y \in K$$

Exercise Prove that for any convex set, there exists a separation oracle. Returning to our examples,

5. All that is needed is to check each constraint and return a violated constraint if it is the case. In the complete formulation, finding a violating inequality can be done in polynomial time but is nontrivial.

6. A hyperplane tangent to the ball of radius R can be easily constructed.

7. Given $\bar{X}$, the separation oracle reduces to finding a vector v such that $\bar{X}v = -\lambda v$ for some positive scalar λ. This would imply that

$$v^T \bar{X} v = \lambda \Rightarrow \bar{X} \cdot (vv^T) = -\lambda,$$

while $X \cdot (vv^T) \geq 0$ for all feasible points.

3.1.1 Reducing Optimization to Feasibility

Before proceeding, we will reduce the problem (3.1), under the convexity assumptions previous mentioned, to a feasibility problem. That is, given a convex set K, find a point on K or prove that K is empty ("give a certificate" that K is empty).

For a scalar parameter t, we define the convex set

$$K \cap \{x \in \mathbb{R}^n : f(x) \leq t\}$$

If we can find a point in such a convex set, then we can solve (3.1) via binary search for the optimal t.

3.1.2 Feasibility Problem

Now, we are going to investigate a convex feasibility algorithm with the following rules:

Input: separate oracle for K, and $r, R \in \mathbb{R}_{++}$
(where a ball of radius r contained in K, K is contained in a ball of radius R)

Output: A point in K if K is not empty, or give a certificate that K is empty.

Theorem 9. *Any algorithm needs $n \log_2 \frac{R}{r}$ oracle calls in the worst case.*

Proof. Since we only have access to the oracle, the set K may be in the largest remaining set at each iteration. In this case, we can reduce the total volume at most by $1/2$. It may happen for iteration k as long,

$$\left(\frac{1}{2}\right)^k < \left(\frac{r}{R}\right)^n$$

which implies that

$$k < n \log_2 \frac{R}{r}$$

□

In the seminal work of Khachiyan and its extension by GLS, the Ellipsoid Algorithm was proved to require $\mathcal{O}(n^2 \log \frac{R}{r})$ iterations for the convex feasibility problem. Subsequently, Karmarkar introduced interior points methods for the special case of linear programming; the latter turned out to be more practical (and polynomial-time).

We are going to consider the following algorithm for convex feasibility.

1. $P = \text{cube}(R)$, $z = 0$.
2. Call Separation Oracle for z. If $z \in K$ stop.
3. $P := P \cap \{x \in \mathbb{R}^n : a^T x \leq a^T z\}$
4. Define the new point $z \in P$.
5. Repeat steps 2, 3, and 4 N times
6. Declare that K is empty.

Two of the steps are not completely defined. It is intuitive that the choice of the points z will define the number of iterations needed, N. Many choices are possible here, e.g., the centroid, a random point, the average of random points, analytic center, point at maximum distance from the boundary, etc.

For instance, consider the centroid and an arbitrary hyperplane crossing it. As we proved in the previous lecture, P will be divided into two convex bodies, each having at least $\frac{1}{e}$ of the total volume of the starting set (Grunbaum's theorem). Since the initial volume is R^n and the final volume is at least r^n, we need at most

$$\log_{1-\frac{1}{e}} \left(\frac{R}{r}\right)^n = \mathcal{O}\left(n \log \frac{R}{r}\right)$$

iterations.

So, we have reduced convex optimization to convex feasibility, and the latter we reduced to finding centroid. Unfortunately, the problem of finding the centroid is #P-hard.

Next, let us consider z to be defined as a random point from P. For simplicity, consider the case where P is a ball of radius R centered at the origin. We would like z to be close to the center, so it is of interest to estimate its expected norm.

$$\mathsf{E}[\|x\|] = \int_0^R \frac{\mathsf{Vol}(S_n)t^{n-1}}{\mathsf{Vol}(B_n)R^n} t dt \tag{3.9}$$

$$= \frac{n}{R^n}\int_0^R t^n dt = \frac{n}{n+1}R. \tag{3.10}$$

That is, a random point in the ball is close to the boundary. Now, we proceed to estimate how much the volume is decreasing at each iteration.

Figure 3.1: Sphere cap.

Defining $b = \mathsf{E}[\|x\|] = \frac{n}{n+1}R$, we obtain $a \leq R\sqrt{\frac{2}{n+1}}$. This implies that the volume is falling by

$$\left(\sqrt{\frac{2}{n+1}}\right)^n,$$

which is decreasing exponentially in n. Thus, it will not give us a polynomial-time algorithm.

Now, consider the average of random points,

$$z = \frac{1}{m}\sum_{i=1}^{m} y^i,$$

where y^i's are i.i.d. points in P.

We will prove shortly the following theorem.

Theorem 10. *Let z and $\{y^i\}_{i=1}^m$ be defined as above, we have*

$$\mathsf{E}[\mathsf{Vol}(P')] \leq \left(1 - \frac{1}{e} + \sqrt{\frac{n}{m}}\right)\mathsf{Vol}(P).$$

Thus, the following corollary will hold.

Corollary 11. *With $m = 10n$, the algorithm finishes in $\mathcal{O}(n \log \frac{R}{r})$ iterations with high probability.*

We will begin to prove Theorem (10) for balls centered at the origin. This implies that $\mathsf{E}[z] = 0$ and, using the radial symmetry,

$$\begin{aligned}
\mathsf{var}(\|z\|) \leq \mathsf{E}[\|z\|^2] &= \frac{1}{m}\mathsf{E}[|y_i|^2] \\
&= \frac{1}{m}\int_0^R \frac{\mathsf{Vol}(S_n)t^{n-1}}{\mathsf{Vol}(B_n)R^n}t^2 dt \\
&= \frac{n}{mR^n}\int_0^R t^{n+1}dt = \frac{n}{m(n+2)}R^2.
\end{aligned}$$

To conclude the proof for this special case, using the inequality

$$P\left(\|z\| > \frac{R}{\sqrt{n}}\right) \leq \frac{Var(z)n}{R^2} = \frac{nR^2}{m(n+2)}\frac{n}{R^2} \leq \frac{n}{3m}, \tag{3.11}$$

using $m = \mathcal{O}(n)$, with a fixed probability, we obtain $\|z\| \leq \frac{R}{\sqrt{n}}$. Now $b = \frac{R}{\sqrt{n}}$, thus,

$$a^2 \quad = \left(R + \tfrac{R}{\sqrt{n}}\right)\left(R - \tfrac{R}{\sqrt{n}}\right) = R^2\left(1 - \tfrac{1}{n}\right). \tag{3.12}$$

This implies that the volume is falling by

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

To extend the previous result to an ellipsoid E, it is enough to observe that if $A : \mathbb{R}^n \to \mathbb{R}^n$ is an affine linear transformation, $E = AB$, which scales the volume of every subset by the same factor, namely, $|\det(A)|$.

To generalize for arbitrary convex sets, we introduce the following concept.

Definition 12. *A set S is in isotropic position if for a random point $x \in S$,*

1. $\mathsf{E}[x] = 0$;
2. $\forall v \in \mathbb{R}^n$, $\mathsf{E}[(v^T x)^2] = \|v\|^2$.

Next, we are going to prove a lemma which proves an equivalence between the second condition of isotropic position and conditions over the variance-covariance matrix of the random vector x.

Lemma 13. *Let x be a random point in a convex set S, we have $\forall v \in \mathbb{R}^n$, $\mathsf{E}[(v^T x)^2] = \|v\|^2$ if and only if $\mathsf{E}[xx^T] = I$.*

Proof. If $\mathsf{E}[(v^T x)^2] = \|v\|^2$, for $v = e_i$ we have

$$\mathsf{E}[(e_i^T x)^2] = \mathsf{E}[x_i^2] = 1.$$

For $v := (v_1, v_2, 0, \ldots, 0)$,

$$\begin{aligned}\mathsf{E}[(v^T x)^2] &= \mathsf{E}[v_1^2 x_1^2 + v_2^2 x_2^2 + 2v_1 v_2 x_1 x_2] \\ &= v_1^2 + v_2^2 + 2v_1 v_2 \mathsf{E}[x_1 x_2] \\ &= v_1^2 + v_2^2,\end{aligned}$$

where the last equality follows from the second condition of Definition 12. So, $\mathsf{E}[x_1 x_2] = 0$.

To prove the other direction,

$$\mathsf{E}[(v^T x)^2] = \mathsf{E}[(v^T x x^T v)] = v^T \mathsf{E}[xx^T] v = v^T v = \|v\|^2.$$

□

It is trivial to see that the first condition of isotropic position can be achieved by a translation. It turns out that is possible to obtain the second condition by means of a linear transformation.

Lemma 14. *Any convex set can be put in isotropic position by an affine transformation.*

Proof. $\mathsf{E}_K[xx^T] = A$ is positive definite. Then, we can write $A = B^2$ (or BB^T).

Defining, $y = B^{-1}x$, we get

$$\mathsf{E}[yy^T] = B^{-1}\mathsf{E}[xx^T](B^{-1})^T = I.$$

The new set is defined as

$$K' := B^{-1}(K - z(K)),$$

where $z(K) := \mathsf{E}[x]$, the centroid of K. □

Note that the argument used before to relate the proof for balls to ellipsoids states that volume ratios are invariant under affine transformations. Thus, we can restrict ourselves to the case where the convex set is in isotropic position due to the previous lemma.

Lemma 15. *K is a convex body, z is the average of m random points from K. If H is a half space containing z,*

$$\mathsf{E}[\mathsf{Vol}(H \cap K)] \geq \left(\frac{1}{e} - \sqrt{\frac{n}{m}}\right)\mathsf{Vol}(K).$$

Proof. As state before, we can restrict ourselves to the case which K is in isotropic position. Since $z = \frac{1}{m}\sum_{i=1}^{m} y^i$,

$$\begin{aligned}\mathsf{E}[\|z\|^2] = \frac{1}{m^2}\sum_{i=1}^{m}\mathsf{E}[\|y^i\|^2] &= \frac{1}{m}\mathsf{E}[\|y^i\|^2] \\ &= \frac{1}{m}\sum_{j=1}^{n}\mathsf{E}[(y_j^i)^2] = \frac{n}{m},\end{aligned}$$

where the first equality follows from the independence between y^i's, and equalities of the second line follows from the isotropic position.

Let h be a unit vector normal to H, since the isotropic position is invariant under rotations, we can assume that $h = e_1 = (1, 0, \ldots, 0)$.

Define the $(n-1)$-volume of the slice $K \cap \{x \in \mathbb{R}^n : x_1 = y\}$ as $\mathsf{Vol}_{n-1}(K \cap \{x \in \mathbb{R}^n : x_1 = y\})$. We can also define the following density function,

$$f_K(y) = \frac{\mathsf{Vol}_{n-1}(K \cap \{x \in \mathbb{R}^n : x_1 = y\})}{\mathsf{Vol}(K)}. \tag{3.13}$$

Note that this function has the following properties:

1. $f_K(y) \geq 0$ for all $y \in \mathbb{R}$;
2. $\int_{\mathbb{R}} f_K(y)dy = 1$;
3. $\int_{\mathbb{R}} y f_K(y)dy = 0$;
4. $\int_{\mathbb{R}} y^2 f_K(y)dy = 1$;

The following technical lemma will give an useful bound to this density (we defer its proof).

Lemma 16. *If f_K satisfies the previous properties for some convex body K,*

$$\max_y f(y) \leq \frac{n}{n+1}\sqrt{\frac{n}{n+2}} < 1$$

Combining this lemma with the bound on the norm of z,

$$\begin{aligned}
\frac{\mathsf{Vol}(K \cap \{x\mathbb{R}^n : x_1 \geq z_1\}}{\mathsf{Vol}(K)} &= \int_{z_1}^{\infty} f_K(y)dy \\
&= \int_0^{\infty} f_K(y)dy - \int_0^{z_1} f_K(y)dy \\
&\geq \frac{1}{e} - \int_0^{z_1} \max_y f_K(y)dy \\
&\geq \frac{1}{e} - \|z\| \\
&\geq \frac{1}{e} - \sqrt{\frac{n}{m}}.
\end{aligned} \tag{3.14}$$

□

Figure 3.2: Construction to increase the moment of inertia I.

Proof. (of Lemma 16) Let K be a convex body in isotropic position for which $\max_y f_K(y)$ is as large as possible. In Figure 3.1.2, we denote by $y^* \in \arg\max_y f_K(y)$, the centroid is zero, and the set K is the black line.

In Figure 3.1.2, the transformation in part I consists of replacing it with a cone whose base is the cross-section at zero and its apex is on the x_1 axis. The distance of the apex is defined to preserve the volume of part I. Such a transformation can only move mass to the left of the centroid.

The transformation in part II of figure 3.1.2 is to construct the convex hull between the cross sections at zero and y^*. This procedure may lose some mass (never increases). Part III is also replaced by a cone, with base being the cross-section at y^* and the distance of the apex is such that we maintain the total volume. Thus, the transformations in parts II and III move mass only to the right, i.e., away from the centroid.

Defining the moment of inertia[4], as $I(K) = \int_K (y - \mathsf{E}_K[y])^2 f_K(y)dy$, we make the following claim.

Claim 17. *Moving mass away from the centroid of a set can only increase the moment of inertia I.*

Let us define a transformation $C \to C'$, $x_1 \to x_1 + g(x_1)$, where $x_1 g(x_1) \geq 0$.

$$\begin{aligned} I(C) &= Var_C(y) = \mathsf{E}_C[y^2] \\ I(C') &= Var_{C'}(y) = \mathsf{E}_{C'}[y^2] - (\mathsf{E}_{C'}[y])^2 \\ &= \mathsf{E}_C[(y+g(y))^2] - (\mathsf{E}_C[y+g(y)])^2 \\ &= Var_C(y) + Var_C(g(y)) + 2\mathsf{E}_C[yg(y)] \\ &\geq I(C). \end{aligned}$$

In fact, we have strict inequality if g(y) is nonzero on a set of positive measure. In this case, we can shrink the support of f_K by a factor strictly smaller that one, and scale up f_K by a factor greater than one to get a set with moment of inertia equal to 1. In this process $\max f_K$ has increased. This is a contradiction since we started with maximum possible. Thus, K must have the shape of K'.

Now, we proceed for the transformation in Figure 3.1.2.

In Figure 3.1.2, y is chosen to preserve the volume in Part I, while Part II is not altered. Note that all the mass removed from the segment

[4]Note that $\mathsf{E}_K[y] = 0$ since K is in isotropic position.

Figure 3.3: Construction to reduce to a double cone.

between $[-y, y]$ is replaced at a distance to zero greater than y. Thus, the moment of inertia increases. By the same argument used before, we obtain a contradiction. Hence, K must be a double cone.

Once we obtain a double cone, we can move y^* to the last point to the right, and build a cone with the same volume which has a greater moment of inertia. So K must be this cone with length h and with base area A at y^*. Then, it is an easy exercise in integration to get $h = (n+1)\sqrt{\frac{n+2}{n}}$. Since $\mathsf{Vol}(K) = \frac{Ah}{n}$,

$$f(y^*) = \frac{A}{\mathsf{Vol}(K)} = \frac{n}{h} = \frac{n}{n+1}\sqrt{\frac{n}{n+2}} < 1.$$

□

3.1.3 Membership Oracle

We have reduced quasi-convex optimization over convex sets to a feasibility problem and solved the latter using a polynomial number of calls to a separation oracle.

We note that with only a membership oracle, given an arbitrary point x in the space, it will return a certificate that $x \in K$ or a certificate that $x \notin K$, and a point $x \in K$ such that $x + rB \subseteq K \subseteq x + RB$, the optimization problem can still be solved using a polynomial number of oracle calls. The idea is to use the oracle to generate random samples (this part needs only a membership oracle), then restrict the set using the objective function values at the sample points. The first polynomial algorithm in this setting was given by GLS using a variant of the Ellipsoid algorithm.

Chapter 4

Sampling by Random Walks

Here we consider the problem of selecting a point uniformly at random from a convex set K. Before attempting a general procedure, let's consider some special cases where we can readily derive sampling procedures.

Cuboid A cuboid is a product of one dimensional intervals, i.e., a set of the form $[a_1, b_1] \times \cdots \times [a_n, b_n]$. Since it is easy to sample uniformly from an interval, we can pick a point in the cube as $(r_1, \ldots, r_n)$, where r_i is chosen uniformly from $[a_i, b_i]$.

Ball All points at distance r from the center of the ball have an equal probability (density) of being chosen. We can choose a random direction by picking a point according to any spherically symmetric distribution, e.g, by picking each coordinate independently from the standard normal. Once we have chosen a direction, it remains to choose a radius. The density of radius r should be proportional to r^{n-1}. This is a one dimensional distribution, so it is easy to sample.

Ellipsoid Any ellipsoid is some linear transformation of a ball, so we can pick a random point in the ball and then map it into the ellipsoid by the linear transformation.

Simplex Any simplex can be reduced by a linear transformation to the set

$$\{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i \leq 1,\ x_i \geq 0\ \forall i\}.$$

Equivalently, we can write this in one higher dimension, introducing a

slack variable, as the set

$$\{x \in \mathbb{R}^{n+1} \mid \sum_{i=1}^{n+1} x_i = 1,\ x_i \geq 0\ \forall i\}.$$

We can sample this set by choosing n points $d_i \in [0,1]$, sorting the d_is, and then assigning $x_i = d_i - d_{i-1}$, where $d_{n+1} = 1$ and $d_0 = 0$. It is clear that $\sum_{i=1}^{n+1} x_n = \sum_{i=1}^{n+1} d_i - d_{i-1} = d_{n+1} - d_0 = 1$ and that $x_i \geq 0$ since the d_is are sorted. It is a good exercise to prove that this gives a uniform distribution.

Next, let's consider how we would sample from a polytope. Any polytope is contained in some sufficiently large axis-aligned cube. One idea we might try is to pick uniformly from the cube until we get a point in the polytope. How long should we expect to wait before we get a point in the polytope? The expected number of iterations is equal to the volume of the cube divided by the volume of the polytope. To get a lower bound, consider the ball of unit radius fit inside of the cube with side length 2. (The cross polytope, for example, is contained in the ball and just fits in the cube.) The volume of this ball is roughly $1/(n/2)^{n/2}$ as $n \to \infty$, whereas the volume of the cube is 2^n, so we would expect to wait roughly $(2n)^{n/2}$ iterations. As another example, if the polytope were a regular simplex, then the expected number of iterations would be $n!$ by the same argument. These examples show that this method is inefficient.

4.1 Grid Walk

To sample from a general convex body, we will use a **random walk**. The basic idea is as follows. We want to sample from some space of points Ω. Each point $x \in \Omega$ has a set of neighbors $\mathcal{N}(x)$. We are given an initial point $x \in \Omega$. The walk proceeds by choosing a random $y \in \mathcal{N}(x)$, setting $x = y$, and repeating. After sufficiently many iterations, under some conditions, the distribution of the current points converges to the stationary distribution.

Our first attempt will be a discrete random walk (with a finite state space). We will use this to highlight many of the issues that come up. In later lectures, we will develop generalizations for arbitrary Markov chains.

The following algorithm is called the Grid-Walk. It depends on a parameter δ, the grid width, which we will specify later. It assumes that we have a membership oracle for K and that $\mathbb{B}_n \subseteq K \subseteq R\,\mathbb{B}_n$.

Figure 4.1: This shows a cube, in the upper-right, whose grid point is in K but all of whose neighboring grid points are not in K. Thus, the space of grid points in K is not connected.

Algorithm Grid-Walk(δ):

1. Let x be a starting point in K.
2. Repeat sufficiently many times:
   - Choose a random $v \in \{\pm e_1, \ldots, \pm e_n\}$.
   - Let $y = x + \delta v$.
   - If $y \in K$ (oracle call), set $x = y$.

In the Grid-Walk, the neighbors of x are the $2n$ points $x \pm \delta e_i$. One consequence of this is that the final point chosen will not be random from all of K, but only from the regular lattice (grid) of points in K whose coordinates are integral multiples of δ away from those of the starting point. This seems unsatisfactory, so let's consider how we could fix it.

One idea would be to choose a random point in the cube $\prod_i [x_i - \frac{1}{2}\delta, x_i + \frac{1}{2}\delta]$, where x is the final grid point chosen by the method above. However, this would not give us only points of K — it would give some points not in K, and it would miss some points in K. We could try to fix the former problem by picking random points from this cube until we get one in K. Unfortunately, this gives equal probability to all cubes, even those that are not fully contained in K, which is not uniform. To get a uniform distribution, each cube needs to be chosen with probability proportional to the volume of its intersection with K. We can fix this problem by starting over from scratch when the point chosen from the cube is not in K.

This method can also miss points in K because the space of grid points that are in K may not be connected, i.e., there may not be a path through neighbors in K that connects some points. Figure 4.1 provides an example. One idea for fixing this would be to include any grid point whose cube intersects K. But how could we check this for a given grid point, i.e., how could we extend the membership oracle to the set of points whose cube intersects K?

Here is a better idea: we scale K to $K' = (1 + \alpha)K$. Consider any grid point whose cube intersects K. The length of the diagonal of the cube is $\delta\sqrt{n}$. Let y be the closest point in K on the line between the origin and the far corner of the cube, as shown in Figure 4.2. Since $\mathbb{B}_n \subseteq K$, the distance of

Figure 4.2: This shows K, $\mathbb{B}_n$ inside K, and a cube that intersects K but whose grid point is not in K. Since $\|y\| \geq 1$ and the length L of the diagonal of the cube is $\delta\sqrt{n}$, the cube is fully contained in $(1+\delta\sqrt{n})K$.

y from origin is at least 1, so if we set $\alpha = \delta\sqrt{n}$, then $(1+\alpha)K$ will contain all of this cube. Thus, any cube that intersects K will be fully contained in K'. Now, consider the line between any two points $x, y \in K$. There is a path through the cubes intersecting this line that only transitions between cubes that share a common side, i.e., whose grid points are neighbors. Every point on the line is in K, so every cube on the path is in K'. This proves that there is a path between any two grid points in K, so our state space is connected. Another way to include all intersected cubes is to expand K to $K + \delta\sqrt{n}\mathbb{B}_n$. It is always possible to extend the membership oracle for K to one for $K + \delta\sqrt{n}\mathbb{B}_n$.

We now have a procedure that will choose a point from K uniformly at random (provided that the final grid point is uniform — more on this below). However, we may have to choose many random grid points before the point chosen from its cube is in K. The probability of success on each iteration is $\mathsf{Vol}(K)$ divided by the volume of all the δ-cubes whose centers are in K'. By the argument from above, every such cube is contained in $(1+\alpha)K' = (1+\alpha)^2 K$. So the probability of success is at least

$$\frac{\mathsf{Vol}(K)}{\mathsf{Vol}((1+\alpha)^2 K)} = (1+\alpha)^{-2n} = (1+\delta\sqrt{n})^{-2n} = (1+\frac{\delta n}{\sqrt{n}})^{-2n} \approx e^{-2\delta n\sqrt{n}}.$$

If we choose $\delta = 1/2n\sqrt{n}$ (polynomial in n), then the probability of success is bounded from below by a constant, which means that the expected number of iterations is $O(1)$.

It remains to determine the number of steps of the random walk needed to choose a random grid point. Let C denote the set of grid points in K, and let $\{w_i\}$ be the set of points traversed by the random walk, with $w_i \in C$ and w_0 the starting point. Then, for any points $x, y \in C$, the probability of transitioning from x to y, $\mathsf{P}(w_{i+1} = y \mid w_i = x)$, is

$$p_{xy} = \begin{cases} \frac{1}{2n} & \text{if } y \in C \cap \mathcal{N}(x), \\ 1 - \frac{|C \cap \mathcal{N}(x)|}{|\mathcal{N}(x)|} & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $p_{xy} = p_{yx} = 1/2n$ if $x, y \in C$ are neighbors. Now, suppose that π_x were a steady state of this distribution, i.e., that the distribution

of w_{i+1} is π if the distribution of w_i is π. For the uniform distribution, we get $\pi_x = \sum_y \pi_y p_{yx} = \pi_x(\sum_y p_{yx}) = \pi_x(\sum_y p_{xy}) = \pi_x$, which shows that the uniform distribution is a steady state. This is good, but this doesn't prove that the random walk converges to this steady state. Perhaps there is more than one steady state, or perhaps it doesn't converge at all.

Suppose that we always performed a transition (i.e, $p_{xx} = 0$ for all $x \in C$). Then, the transition graph would be bipartite (since the grid is bipartite), so the random walk would clearly not be able to converge to the uniform distribution. We can see that our graph is bipartite except when we get to boundary points. So the distribution of points that could be reached by the random walk could not be uniform until we have a reasonable probability of getting to boundary points. It turns out that this works alright for convex bodies. However, a more convenient approach is to modify the distribution so that we always stay at point x with probability at least $\frac{1}{2}$: on every step, we first flip a coin to decide if we will try to move at all. This destroys the bipartition of the interior grid points. (In particular, the transition matrix P becomes positive semidefinite, so all eigenvalues are nonnegative. The back-and-forth transitions of a bipartition give a negative eigenvalue, which is then ruled out.) A random walk that stays in place with probability $\frac{1}{2}$ and attempt the walk with probability $\frac{1}{2}$ is called **lazy**.

Another potential difficulty for a random walk is having a disconnected state space. We saw above that expanding K took care of our connectedness problems.

Now, let's return to the question of whether the random walk converges to the uniform distribution. This is answered by the following classical theorem, whose proof uses the Perron-Frobenius theorem.

Theorem 18. *If a transition matrix P is nonbipartite and irreducible, then the random walk tends to a unique steady state.*

This does not answer the question of *how fast* the random walk tends to the uniform distribution. So rather than proving this, we will prove a stronger theorem that bounds the speed of convergence. For this theorem, we will need the following notation. Let Q_i be the distribution of point w_i of the random walk. We define the distance between Q_i and the steady state π to be $|Q_i - \pi| = \frac{1}{2}\sum_{x\in\Omega} |Q_i(x) - \pi(x)|$. The factor of $\frac{1}{2}$ makes 1 the maximum value (this is achieved when the two distributions are disjoint).

We next define an important notion that will help bound the rate of convergence to the steady state.

Definition 19. *For any $S \subset \Omega$, let*

$$\Phi(S) = \frac{\mathsf{P}(w_{i+1} \in \bar{S} \mid w_i \in S)}{\pi(S)}.$$

Then, the conductance of Ω is

$$\phi = \min_{S \subset \Omega} \Phi(S) = \min_{S \subset \Omega,\, \pi(S) \leq \frac{1}{2}} \Phi(S).$$

Theorem 20 (Jerrum-Sinclair)**.** *Suppose we start at a discrete random walk at a point x Then,*

$$|Q_t - \pi| \leq \frac{1}{Q(t)} \left(1 - \frac{\phi^2}{2}\right)^t.$$

This theorem implies that, the "distance" to the stationary distribution falls by a constant factor in $O(1/\phi^2)$ steps. For this reason, the quantity $1/\phi^2$ is often called the **mixing time**. On the other hand, it is not hard to see that $1/\phi$ is a lower bound on the number of iterations needed to decrease the distance by a constant factor since this is the expected number of iterations required to cross the bottleneck, from S to $\bar{S}$ (consider a starting distribution Q_0 with probability 1 at $w_0 \in S$).

From this theorem, we can see that the random walk will converge rapidly to a uniform distribution when the state space has high conductance. Intuitively, the grid points in a convex sets should have a good chance of having high conductance since they avoid the bottlenecks associated with concavity. Before exploring this further, we will generalize Theorem 20 to arbitrary Markov chains.

Chapter 5

Convergence of Markov Chains

A random walk is a Markov chain. We have a state space Ω and a set of subsets $\mathcal{A}$ of Ω that form a σ-algebra, i.e., $\mathcal{A}$ is closed under complements and countable unions. It is an immediate consequence that $\mathcal{A}$ is closed under countable intersections and that $\emptyset, \Omega \in \mathcal{A}$. For any $u \in \Omega$ and $A \in \mathcal{A}$, we have the one-step probability, $P_u(A)$, which tells us the probability of being in A after taking one step from u. Lastly, we have a starting distribution Q_0 on Ω, which gives us a probability $Q_0(A)$ of starting in the set $A \in \mathcal{A}$.

With this setup, a **Markov chain** is a sequence of points $w_0, w_1, w_2, \ldots$ such that $\mathsf{P}(w_0 \in A) = Q_0(A)$ and

$$\mathsf{P}(w_{i+1} \in A \mid w_0 = u_0, \ldots, w_i = u_i) = \mathsf{P}(w_{i+1} \in A \mid w_i = u_i) = P_{u_i}(A),$$

for any $A \in \mathcal{A}$. A distribution Q is called **stationary** if, for all $A \in \mathcal{A}$,

$$Q(A) = \int_{u \in \Omega} P_u(A)\ dQ(u).$$

In other words, Q is stationary if the probability of being in A is the same after one step. We also have a generalized version of the symmetry we saw above in the transition probabilities ($p_{xy} = p_{yx}$). The Markov chain is called **time-reversible** if, for all $A, B \in \mathcal{A}$,

$$\mathsf{P}(w_{i+1} \in B \mid w_i \in A) = \mathsf{P}(w_{i+1} \in A \mid w_i \in B).$$

5.1 Example: The ball walk

The following algorithm, called Ball-Walk, is a continuous random walk. In this case, the set corresponding to the neighborhood of x is $B_\delta(x) = x + \delta\mathbb{B}_n$.

Algorithm Ball-Walk(δ):

1. Let x be a starting point in K.
2. Repeat sufficiently many times:
   - Choose a random $y \in B_\delta(x)$.
   - If $y \in K$, set $x = y$.

Here, the state space is all of the set, so $\Omega = K$. The σ-algebra is the measurable subsets of K, as is usual. We can define a density function for the probability of transitioning from $u \in K$ to v, provided that $u \neq v$:

$$p(u,v) = \begin{cases} \frac{1}{\mathsf{Vol}(\delta\mathbb{B}_n)} & \text{if } v \in K \cap B_\delta(u), \\ 0 & \text{otherwise.} \end{cases}$$

The probability of staying at u is

$$P_u(u) = 1 - \frac{\mathsf{Vol}(K \cap B_\delta(u))}{\mathsf{Vol}(\delta\mathbb{B}_n)}.$$

Putting these together, the probability of transitioning from u to any measurable subset A is

$$P_u(A) = \begin{cases} \frac{\mathsf{Vol}(A \cap K \cap B_\delta(u))}{\mathsf{Vol}(\delta\mathbb{B}_n)} + P_u(u) & \text{if } u \in A \\ \frac{\mathsf{Vol}(A \cap K \cap B_\delta(u))}{\mathsf{Vol}(\delta\mathbb{B}_n)} & \text{if } u \notin A. \end{cases}$$

Since the density function is symmetric, it is easy to see that the uniform distribution is stationary. We can also verify this directly. We can compute the probability of being in A after one step by adding up the probability that we transition to u for each $u \in A$. Thus, after one step from the uniform distribution $dQ(u) = \frac{1}{\mathsf{Vol}(K)}\, du$, the probability of being in A is

$$\begin{aligned} &\textstyle\int_{u\in A} P_u(u)\, dQ(u) + \int_{u\in A} \int_{v\in K\cap B_\delta(u)\setminus\{u\}} \frac{1}{\mathsf{Vol}(\delta\mathbb{B}_n)}\, dQ(v)\, du \\ &= \textstyle\int_{u\in A} \left(1 - \frac{\mathsf{Vol}(K\cap B_\delta(u))}{\mathsf{Vol}(\delta\mathbb{B}_n)}\right)\, dQ(u) + \int_{u\in A} \int_{v\in K\cap B_\delta(u)} \frac{1}{\mathsf{Vol}(\delta\mathbb{B}_n)}\, dQ(v)\, du \\ &= \textstyle\int_{u\in A}\, dQ(u) + \frac{1}{\mathsf{Vol}(\delta\mathbb{B}_n)} \int_{u\in A} \left(\int_{v\in K\cap B_\delta(u)}\, dv - \mathsf{Vol}(K\cap B_\delta(u))\right)\, dQ(u) \\ &= \textstyle\int_{u\in A}\, dQ(u) \\ &= Q(A) \end{aligned}$$

5.2 Ergodic flow and conductance

Another way to look at the one-step probability distributions is in terms of flow. For any subset $A \in \mathcal{A}$, the ergodic flow $\Phi(A)$ is the probability of transitioning from A to $\Omega \setminus A$, i.e.,

$$\Phi(A) = \int_{u \in A} P_u(\Omega \setminus A)\ dQ.$$

Intuitively, in a stationary distribution, we should have $\Phi(A) = \Phi(\Omega \setminus A)$. In fact, this is a characterization of the stationary distribution.

Theorem 21. *A distribution Q is stationary iff $\Phi(A) = \Phi(\Omega \setminus A)$ for all $A \in \mathcal{A}$.*

Proof. Consider their difference

$$\begin{aligned} \Phi(A) - \Phi(\Omega \setminus A) &= \textstyle\int_{u \in A} P_u(\Omega \setminus A)\ dQ(u) - \int_{u \notin A} P_u(A)\ dQ(u) \\ &= \textstyle\int_{u \in A} (1 - P_u(A))\ dQ(u) - \int_{u \notin A} P_u(A)\ dQ(u) \\ &= Q(A) - \textstyle\int_{u \in \Omega} P_u(A)\ dQ(u). \end{aligned}$$

The latter quantity is the probability of staying in A after one step, so $\Phi(A) - \Phi(\Omega \setminus A) = 0$ iff Q is stationary. □

Now, let's return to the question of how quickly (and whether) this random walk converges to the stationary distribution. As before, it is convenient to make the walk lazy by giving probability $\frac{1}{2}$ of staying in place instead of taking a step. This means that $P_u(\{u\}) \geq \frac{1}{2}$ and, more generally, for any $A \in \mathcal{A}$, $P_u(A) \geq \frac{1}{2}$ if $u \in A$ and $P_u(A) < \frac{1}{2}$ if $u \notin A$. Also as before, the notion of conductance will be useful. In this general context, we define conductance as

$$\phi = \min_{A \in \mathcal{A},\ Q(A) \leq \frac{1}{2}} \frac{\Phi(A)}{Q(A)},$$

where Q is the stationary distribution. This is the probability of transitioning to $\Omega \setminus A$ given that we are starting in A.

We use the L_1 distance or total variation distance as the definition of the distance $d(Q_t, Q)$ between distributions:

$$d(Q_t, Q) : |Q_t - Q| = \sup_{A \in \mathcal{A}} Q_t(A) - Q(A).$$

We would like to know how large t must be before $d(Q_t, Q) \leq \frac{1}{2} d(Q_0, Q)$. This is one definition of the mixing time. It would be nice to get a bound

by showing that $Q_t(A) - Q(A)$ drops quickly for every A. This is not the case (consider, e.g., the ball walk starting at a single point). Instead, we can look at $\sup_{Q(A)=x} Q_t(A) - Q(A)$ for each fixed $x \in [0,1]$. A bound for every x would imply the bound we need. To prove a bound on this by induction, we will define it in a formally weaker way. Our upper bound is

$$h_t(x) = \sup_{g \in \mathcal{F}_x} \int_{u \in \Omega} g(u)\ (dQ_t(u) - dQ(u)) = \sup_{g \in \mathcal{F}_x} \int_{u \in \Omega} g(u)\ dQ_t(u) - x,$$

where $\mathcal{F}_x$ is the set of functions

$$\mathcal{F}_x = \left\{ g : \Omega \to [0,1] \ : \ \int_{u \in \Omega} g(u)\ dQ(u) = x \right\}.$$

It is clear that $h_t(x)$ is an upper bound on $\sup_{Q(A)=x} Q_t(A) - Q(A)$ since g could be the characteristic of A. The following lemma shows that these two quantities are in fact equal as long as Q is atom-free, i.e., there is no $x \in \Omega$ such that $Q(\{x\}) > 0$.

Lemma 22. *If Q is atom-free, then $h_t(x) = \sup_{Q(A)=x} Q_t(A) - Q(A)$.*

Proof (Sketch). Consider the set of points X that maximize dQ_t/dQ, their value density. (This part would not be possible of Q had an atom.) Put $g(x) = 1$ for all $x \in X$. These points give the maximum payoff per unit of weight from g, so it is optimal to put as much weight on them as possible. Now, find the set of maximizing points in $\Omega \setminus X$. Set $g(x) = 1$ at these points. Continue until the set of points with $g(x) = 1$ has measure x. □

In fact, this argument shows that when Q is atom-free, we can find a set A that achieves the supremum. When Q has atoms, we can include the high value atoms and use this procedure on the non-atom subset of Ω; however, we made to include a fraction of one atom to achieve the supremum. This shows that $h_t(x)$ can be achieved by a function g that is $0-1$ valued everywhere except for at most one point.

Another important fact about h_t is the following.

Lemma 23. *The function h_t is concave.*

Proof. Let $g_1 \in \mathcal{F}_x$ and $g_2 \in \mathcal{F}_y$. Then, we can see that $\alpha g_1 + (1-\alpha) g_2 \in F_{\alpha x + (1-\alpha)y}$, which implies that

$$\begin{aligned} & h_t(\alpha x + (1-\alpha)y) \\ &\geq \int_{u \in \Omega} \alpha g_1(u) + (1-\alpha) g_2(u)\ dQ_t(u) - (\alpha x + (1-\alpha)y) \\ &= \alpha \left(\int_{u \in \Omega} g_1(u)\ dQ_t(u) - x \right) + (1-\alpha) \left(\int_{u \in \Omega} g_2(u)\ dQ_t(u) - y \right). \end{aligned}$$

Since this holds for any such g_1 and g_2, it must hold if we take the sup, which gives us $h_t(\alpha x + (1-\alpha)y) \geq \alpha h_t(x) + (1-\alpha)h_t(y)$. Thus, h_t is concave. $\square$

Now, we come to the main lemma relating h_t to h_{t-1}. This will allow us to put a bound on $d(Q_t, Q)$.

Lemma 24. *Let Q be atom-free, and $y = \min\{x, 1-x\}$. Then*

$$h_t(x) \leq \frac{1}{2}h_{t-1}(x - 2\phi y) + \frac{1}{2}h_{t-1}(x + 2\phi y).$$

Proof. Assume that $y = x$, i.e., $x \leq \frac{1}{2}$. The other part is similar. We will construct two functions, g_1 and g_2, and use these to bound $h_t(x)$. Let $A \in \mathcal{A}$ be a subset to be chosen later with $Q(A) = x$. Let

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

First, note that $(\frac{1}{2}g_1 + \frac{1}{2}g_2)(u) = P_u(A)$ for all $u \in \Omega$, which means that

$$\begin{aligned} &\frac{1}{2}\int_{u\in\Omega} g_1(u)\ dQ_{t-1}(u) + \frac{1}{2}\int_{u\in\Omega} g_2(u)\ dQ_{t-1}(u) \\ &= \int_{u\in\Omega} (\frac{1}{2}g_1(u) + \frac{1}{2}g_2(u))\ dQ_{t-1}(u) \\ &= \int_{u\in\Omega} P_u(A)\ dQ_{t-1}(u) \\ &= Q_t(A) \end{aligned}$$

On the other hand,

$$\frac{1}{2}\int_{u\in\Omega} g_1(u)\ dQ(u) + \frac{1}{2}\int_{u\in\Omega} g_2(u)\ dQ(u) = \int_{u\in\Omega} P_u(A)\ dQ(u) = Q(A)$$

since Q is stationary. Hence $(\frac{1}{2}g_1 + \frac{1}{2}g_2) \in \mathcal{F}_x$.Putting these together, we have

$$\begin{aligned} &\frac{1}{2}\int_{u\in\Omega} g_1(u)\ (dQ_{t-1}(u) - dQ(u)) + \frac{1}{2}\int_{u\in\Omega} g_2(u)\ (dQ_{t-1}(u) - dQ(u)) \\ &= Q_t(A) - Q(A). \end{aligned}$$

Figure 5.1: The value of $h_t(x)$ lies below the line between $h_{t-1}(x_1)$ and $h_{t-1}(x_2)$. Since $x_1 \leq x(1-2\phi) \leq x \leq x(1+2\phi) \leq x_2$ and h_{t-1} is concave, this implies that $h_t(x)$ lies below the line between $h_{t-1}(x(1-2\phi))$ and $h_{t-1}(x(1+2\phi))$.

Since Q is atom-free, there is a subset $A \subseteq \Omega$ such that

$$\begin{aligned} h_t(x) &= Q_t(A) - Q(A) \\ &= \frac{1}{2}\int_{u\in\Omega} g_1(u)\ (dQ_{t-1}(u) - dQ(u)) + \\ &\quad \frac{1}{2}\int_{u\in\Omega} g_2(u)\ (dQ_{t-1}(u) - dQ(u)) \\ &\leq \frac{1}{2}h_{t-1}(x_1) + \frac{1}{2}h_{t-1}(x_2), \end{aligned}$$

where $x_1 = \int_{u\in\Omega} g_1(u)\ dQ(u)$ and $x_2 = \int_{u\in\Omega} g_2(u)\ dQ(u)$. Now, we know that $\frac{1}{2}x_1 + \frac{1}{2}x_2 = x$. Specifically, we can see that

$$\begin{aligned} x_1 &= \int_{u\in\Omega} g_1(u)\ dQ(u) \\ &= 2\int_{u\in A} P_u(A)\ dQ(u) - \int_{u\in A} dQ(u) \\ &= 2\int_{u\in A} (1 - P_u(\Omega \setminus A))\ dQ(u) - x \\ &= x - 2\int_{u\in A} P_u(\Omega \setminus A)\ dQ(u) \\ &= x - 2\Phi(A) \\ &\leq x - 2\phi x \\ &= x(1-2\phi). \end{aligned}$$

This implies that $x_2 \geq x(1+2\phi)$. Since h_{t-1} is concave, the chord from x_1 to x_2 on h_{t-1} lies below the chord from $x(1-2\phi)$ to $x(1+2\phi)$. (See Figure 5.1.) Therefore, $h_t(x) \leq \frac{1}{2}h_{t-1}(x(1-2\phi)) + \frac{1}{2}h_{t-1}(x(1+2\phi))$. □

Given some information about Q_0, this lemma allows us to put bounds on the rate of convergence to the stationary distribution.

Theorem 25. *Let $0 \leq s \leq 1$ and C_0 and C_1 be such that*

$$h_0(x) \leq C_0 + C_1 \min\{\sqrt{x-s}, \sqrt{1-x-s}\}.$$

Then

$$h_t(x) \leq C_0 + C_1 \min\{\sqrt{x-s}, \sqrt{1-x-s}\} \left(1 - \frac{\phi^2}{2}\right)^t.$$

‘

Proof. We argue by induction on t for $s = 0$. The inequality is true for $t = 0$ by our hypothesis. Now, suppose that $x \leq \frac{1}{2}$ and the inequality holds for all values less than t. From the lemma, we know that

$$\begin{aligned} h_t(x) &\leq \frac{1}{2}h_{t-1}(x(1-2\phi)) + \frac{1}{2}h_{t-1}(x(1+2\phi)) \\ &\leq C_0 + \frac{1}{2}C_1(\sqrt{x(1-2\phi)} + \sqrt{x(1+2\phi)})(1 - \frac{1}{2}\phi^2)^{t-1} \\ &= C_0 + \frac{1}{2}C_1\sqrt{x}(\sqrt{1-2\phi} + \sqrt{1+2\phi})(1 - \frac{1}{2}\phi^2)^{t-1} \\ &\leq C_0 + \frac{1}{2}C_1\sqrt{x}(1 - \frac{1}{2}\phi^2)^t, \end{aligned}$$

provided that $\sqrt{1-2\phi} + \sqrt{1+2\phi} \leq 2(1 - \frac{1}{2}\phi^2)$. The latter inequality can be verified by squaring both sides, rearranging, and squaring again. □

Corollary 26. *If $M = \sup_{A\in\mathcal{A}} Q_0(A)/Q(A)$, then we have*

$$d(Q_t, Q) \leq \sqrt{M}(1 - \tfrac{1}{2}\phi^2)^t.$$

Proof. By the definition of M, we know that

$$h_0(x) \leq \min\{Mx, 1\} - x \leq \min\{Mx, 1\}.$$

Next, we will show that $\min\{Mx, 1\} \leq \sqrt{Mx}$. If $Mx = \min\{Mx, 1\}$, then $Mx \leq 1$, which implies that $Mx \leq \sqrt{Mx}$. If $1 = \min\{Mx, 1\}$, then $1 \leq Mx$, which implies that $1 \leq \sqrt{Mx} \leq Mx$. So we have shown that

$$h_0(x) \leq \min\{Mx, 1\} \leq \sqrt{Mx}.$$

Thus, by the last theorem, we know that

$$d(Q_t, Q) \leq \max_x h_t(x) \leq \sqrt{M}(1 - \tfrac{1}{2}\phi^2)^t.$$

□

Chapter 6

Sampling with the Ball Walk

A geometric random walk is said to be *rapidly mixing* if its conductance is bounded from below by an inverse polynomial in the dimension. By Corollary 26, this implies that the number of steps to halve the variation distance to stationary is a polynomial in the dimension. The conductance of the ball walk in a convex body K can be exponentially small. Consider, for example, starting at point x near the apex of a rotational cone in $\mathbb{R}^n$. Most points in a ball of radius δ around x will lie outside the cone (if x is sufficiently close to the apex) and so the local conductance is arbitrarily small. So, strictly speaking, the ball walk is not rapidly mixing.

There are two ways to get around this. For the purpose of sampling uniformly from K, one can expand K a little bit by considering $K' = K + \alpha B_n$, i.e., adding a ball of radius α around every point in K. Then for $\alpha > 2\delta\sqrt{n}$, it is not hard to see that $\ell(u)$ is at least $1/8$ for every point $u \in K'$. We can now consider the ball walk in K'. This fix comes at a price. First, we need a membership oracle for K'. This can be constructed as follows: given a point $x \in \mathbb{R}^n$, we find a point $y \in K$ such that $|x - y|$ is minimum. This is a convex program and can be solved using the ellipsoid algorithm [**?**] and the membership oracle for K, Second, we need to ensure that $\mathsf{Vol}(K')$ is comparable to $\mathsf{Vol}(K)$. Since K contains a unit ball, $K' \subseteq (1+\alpha)K$ and so with $\alpha < 1/n$, we get that $\mathsf{Vol}(K') < e\mathsf{Vol}(K)$. Thus, we would need $\delta < 1/2n\sqrt{n}$.

Does large local conductance imply that the conductance is also large? We will prove that the answer is yes. The next lemma about one-step distributions of nearby points will be useful.

Lemma 27. *Let u, v be such that $|u - v| \leq \frac{t\delta}{\sqrt{n}}$ and $\ell(u), \ell(v) \geq \ell$. Then,*

$$||P_u - P_v||_{tv} \leq 1 + t - \ell.$$

Roughly speaking, the lemma says that if two points with high local conductance are close in Euclidean distance, then their one-step distributions have a large overlap. Its proof follows from a computation of the overlap volume of the balls of radius δ around u and v.

We can now state and prove a bound on the conductance of the ball walk.

Theorem 28. *Let K be a convex body of diameter D so that for every point u in K, the local conductance of the ball walk with δ steps is at least ℓ. Then,*

$$\phi \geq \frac{\ell^2 \delta}{16\sqrt{n}D}.$$

The structure of most proofs of conductance is similar and we will illustrate it by proving this theorem.

Proof. Let $K = S_1 \cup S_2$ be a partition into measurable sets. We will prove that

$$\int_{S_1} P_x(S_2)\, dx \quad \geq \quad \frac{\ell^2 \delta}{16\sqrt{n}D} \min\{\mathsf{Vol}(S_1), \mathsf{Vol}(S_2)\} \tag{6.1}$$

Note that since the uniform distribution is stationary,

$$\int_{S_1} P_x(S_2)\, dx = \int_{S_2} P_x(S_1)\, dx.$$

Consider the points that are "deep" inside these sets, i.e. unlikely to jump out of the set (see Figure 6.1):

$$S_1' = \left\{ x \in S_1 : P_x(S_2) < \frac{\ell}{4} \right\}$$

and

$$S_2' = \left\{ x \in S_2 : P_x(S_1) < \frac{\ell}{4} \right\}.$$

Let S_3' be the rest i.e., $S_3' = K \setminus S_1' \setminus S_2'$.

Figure 6.1: The conductance proof. The dark line is the boundary between S_1 and S_2.

Suppose $\mathsf{Vol}(S_1') < \mathsf{Vol}(S_1)/2$. Then

$$\int_{S_1} P_x(S_2)\,dx \geq \frac{\ell}{4}\mathsf{Vol}(S_1 \setminus S_1') \geq \frac{\ell}{8}\mathsf{Vol}(S_1)$$

which proves (6.1).

So we can assume that $\mathsf{Vol}(S_1') \geq \mathsf{Vol}(S_1)/2$ and similarly $\mathsf{Vol}(S_2') \geq \mathsf{Vol}(S_2)/2$. Now, for any $u \in S_1'$ and $v \in S_2'$,

$$||P_u - P_v||_{tv} \geq 1 - P_u(S_2) - P_v(S_1) > 1 - \frac{\ell}{2}.$$

Applying Lemma 27 with $t = \ell/2$, we get that

$$|u - v| \geq \frac{\ell\delta}{2\sqrt{n}}.$$

Thus $d(S_1, S_2) \geq \ell\delta/2\sqrt{n}$. Applying Theorem 29 to the partition S_1', S_2', S_3', we have

$$\begin{aligned}
\mathsf{Vol}(S_3') &\geq \frac{\ell\delta}{\sqrt{n}D}\min\{\mathsf{Vol}(S_1'), \mathsf{Vol}(S_2')\} \\
&\geq \frac{\ell\delta}{2\sqrt{n}D}\min\{\mathsf{Vol}(S_1), \mathsf{Vol}(S_2)\}
\end{aligned}$$

We can now prove (6.1) as follows:

$$\begin{aligned}
\int_{S_1} P_x(S_2)\,dx &= \frac{1}{2}\int_{S_1} P_x(S_2)\,dx + \frac{1}{2}\int_{S_2} P_x(S_1)\,dx \\
&\geq \frac{1}{2}\mathsf{Vol}(S_3')\frac{\ell}{4} \\
&\geq \frac{\ell^2\delta}{16\sqrt{n}D}\min\{\mathsf{Vol}(S_1), \mathsf{Vol}(S_2)\}.
\end{aligned}$$

□

As observed earlier, by going to $K' = K + (1/n)B_n$ and using $\delta = 1/2n\sqrt{n}$, we have $\ell \geq 1/8$. Thus, for the ball walk in K', $\phi = \Omega(1/n^2D)$ and the mixing rate is $O(n^4D^2)$.

In later chapters, we will see how to improve this bound. We will also see how to obtain a polynomial sampling algorithm, i.e., one whose dependence on D is only logarithmic in D.

Chapter 7

The Localization Lemma and an Isoperimetric Inequality

To formulate an isoperimetric inequality for convex bodies, we consider a partition of a convex body K into three sets S_1, S_2, S_3 such that S_1 and S_2 are "far" from each other, and the inequality bounds the minimum possible volume of S_3 relative to the volumes of S_1 and S_2. We will consider different notions of distance between subsets. Perhaps the most basic is the Euclidean distance:

$$d(S_1, S_2) = \min\{|u - v| \,:\, u \in S_1, v \in S_2\}.$$

Suppose $d(S_1, S_2)$ is large. Does this imply that the volume of $S_3 = K \setminus (S_1 \cup S_2)$ is large? The classic counterexample to such a theorem is a dumbbell — two large subsets separated by very little. Of course, this is not a convex set!

The next theorem, proved in [**?**] (improving on a theorem in [**?**] by a factor of 2; see also [**?**]) asserts that the answer is yes.

Theorem 29. *Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D. Then,*

$$\mathsf{Vol}(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\mathsf{Vol}(S_1), \mathsf{Vol}(S_2)\}.$$

A limiting version of this inequality is the following: For any subset S of a convex body of diameter D,

$$\mathsf{Vol}_{n-1}(\partial S \cap K) \geq \frac{2}{D} \min\{\mathsf{Vol}(S), \mathsf{Vol}(K \setminus S)\}$$

which says that the surface area of S inside K is large compared to the volumes of S and $K \setminus S$. This is in direct analogy with the classical isoperimetric inequality, which says that the surface area to volume ratio of any measurable set is at least the ratio for a ball.

How does one prove such an inequality? In what generality does it hold? (i.e., for what measures besides the uniform measure on a convex set?) We will address these questions in this section.

Theorem 29 can be generalized to arbitrary logconcave measures. Its proof is very similar to that of 29 and we will presently give an outline.

Theorem 30. *Let f be a logconcave function whose support has diameter D and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

In terms of diameter, this inequality is the best possible, as shown by a cylinder. A more refined inequality is obtained in [**?**, **?**] using the average distance of a point to the center of gravity (in place of diameter). It is possible for a convex body to have much larger diameter than average distance to its centroid (e.g., a cone). In such cases, the next theorem provides a better bound.

Theorem 31. *Let f be a logconcave density in $\mathbb{R}^n$ and π_f be the corresponding measure. Let z_f be the centroid of f and define $M(f) = \mathsf{E}_f(|x - z_f|)$. Then, for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{\ln 2}{M(f)} d(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

For an isotropic density, $M(f)^2 \leq \mathsf{E}_f(|x - z_f|^2) = n$ and so $M(f) \leq \sqrt{n}$. The diameter could be unbounded (e.g., an isotropic Gaussian).

A further refinement, based on isotropic position, has been conjectured in [**?**]. Let λ be the largest eigenvalue of the inertia matrix of f, i.e.,

$$\lambda = \max_{v:|v|=1} \int_{\mathbb{R}^n} f(x)(v^T x)^2 \, dx. \tag{7.1}$$

Then the conjecture says that there is an absolute constant c such that

$$\pi_f(S_3) \geq \frac{c}{\sqrt{\lambda}} d(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

7.0.1 The localization lemma

The proofs of these inequalities are based on an elegant idea: integral inequalities in $\mathbb{R}^n$ can be reduced to one-dimensional inequalities! Checking the latter can be tedious but is relatively easy. We illustrate the main idea by sketching the proof of Theorem 30.

For a proof of Theorem 30 by contradiction, let us assume the converse of its conclusion, i.e., for some partition S_1, S_2, S_3 of $\mathbb{R}^n$ and logconcave density f,

$$\int_{S_3} f(x)\,dx < C\int_{S_1} f(x)\,dx \quad \text{and} \quad \int_{S_3} f(x)\,dx < C\int_{S_2} f(x)\,dx$$

where $C = 2d(S_1, S_2)/D$. This can be reformulated as

$$\int_{\mathbb{R}^n} g(x)\,dx > 0 \quad \text{and} \quad \int_{\mathbb{R}^n} h(x)\,dx > 0 \tag{7.2}$$

where

$$g(x) = \begin{cases} Cf(x) & \text{if } x \in S_1, \\ 0 & \text{if } x \in S_2, \\ -f(x) & \text{if } x \in S_3. \end{cases} \quad \text{and} \quad h(x) = \begin{cases} 0 & \text{if } x \in S_1, \\ Cf(x) & \text{if } x \in S_2, \\ -f(x) & \text{if } x \in S_3. \end{cases}$$

These inequalities are for functions in $\mathbb{R}^n$. The main tool to deal with them is the *localization lemma* [**?**] (see also [**?**] for extensions and applications).

Lemma 32. *Let $g, h : \mathbb{R}^n \to \mathbb{R}$ be lower semi-continuous integrable functions such that*

$$\int_{\mathbb{R}^n} g(x)\,dx > 0 \quad \textit{and} \quad \int_{\mathbb{R}^n} h(x)\,dx > 0.$$

Then there exist two points $a, b \in \mathbb{R}^n$ and a linear function $\ell : [0,1] \to \mathbb{R}_+$ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a + tb)\,dt > 0 \quad \textit{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a + tb)\,dt > 0.$$

The points a, b represent an interval A and one may think of $l(t)^{n-1}dA$ as the cross-sectional area of an infinitesimal cone with base area dA. The lemma says that over this cone truncated at a and b, the integrals of g and h are positive. Also, without loss of generality, we can assume that a, b are in the union of the supports of g and h.

The main idea behind the lemma is the following. Let H be any halfspace such that

$$\int_H g(x)\,dx = \frac{1}{2}\int_{\mathbb{R}^n} g(x)\,dx.$$

Let us call this a bisecting halfspace. Now either

$$\int_H h(x)\,dx > 0 \quad \text{or} \quad \int_{R^n\setminus H} h(x)\,dx > 0.$$

Thus, either H or its complementary halfspace will have positive integrals for both g and h. Thus we have reduced the domains of the integrals from $\mathbb{R}^n$ to a halfspace. If we could repeat this, we might hope to reduce the dimensionality of the domain. But do there even exist bisecting halfspaces? In fact, they are aplenty: for any $(n-2)$-dimensional affine subspace, there is a bisecting halfspace with A contained in its bounding hyperplane. To see this, let H be halfspace containing A in its boundary. Rotating H about A we get a family of halfspaces with the same property. This family includes H', the complementary halfspace of H. Now the function $\int_H g - \int_{R^n\setminus H} g$ switches sign from H to H'. Since this is a continuous family, there must be a halfspace for which the function is zero, which is exactly what we want (this is sometimes called the "ham sandwich" theorem).

If we now take all $(n-2)$-dimensional affine subspaces given by some $x_i = r_1$, $x_j = r_2$ where r_1, r_2 are rational, then the intersection of all the corresponding bisecting halfspaces is a line (by choosing only rational values for x_i, we are considering a countable intersection). As long as we are left with a two or higher dimensional set, there is a point in its interior with at least two coordinates that are rational, say $x_1 = r_1$ and $x_2 = r_2$. But then there is a bisecting halfspace H that contains the affine subspace given by $x_1 = r_1, x_2 = r_2$ in its boundary, and so it properly partitions the current set. With some additional work, this leads to the existence of a concave function on an interval (in place of the linear function ℓ in the theorem) with positive integrals. Simplifying further from concave to linear takes quite a bit of work.

Going back to the proof sketch of Theorem 30, we can apply the lemma to get an interval $[a, b]$ and a linear function ℓ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a + tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a + tb)\,dt > 0. \tag{7.3}$$

(The functions g, h as we have defined them are not lower semi-continuous. However, this can be easily achieved by expanding S_1 and S_2 slightly so as to

make them open sets, and making the support of f an open set. Since we are proving strict inequalities, we do not lose anything by these modifications).

Let us partition $[0,1]$ into Z_1, Z_2, Z_3.

$$Z_i = \{t \in [0,1] \,:\, (1-t)a + tb \in S_i\}.$$

Note that for any pair of points $u \in Z_1, v \in Z_2$, $|u-v| \geq d(S_1,S_2)/D$. We can rewrite (7.3) as

$$\int_{Z_3} \ell(t)^{n-1} f((1-t)a+tb)\,dt < C \int_{Z_1} \ell(t)^{n-1} f((1-t)a+tb)\,dt$$

and

$$\int_{Z_3} \ell(t)^{n-1} f((1-t)a+tb)\,dt < C \int_{Z_2} \ell(t)^{n-1} f((1-t)a+tb)\,dt.$$

The functions f and $\ell(.)^{n-1}$ are both logconcave, so $F(t) = \ell(t)^{n-1} f((1-t)a+tb)$ is also logconcave. We get,

$$\int_{Z_3} F(t)\,dt < C \min\left\{ \int_{Z_1} F(t)\,dt, \int_{Z_2} F(t)\,dt \right\}. \tag{7.4}$$

Now consider what Theorem 30 asserts for the function $F(t)$ over the interval $[0,1]$ and the partition Z_1, Z_2, Z_3:

$$\int_{Z_3} F(t)\,dt \geq 2d(Z_1,Z_2) \min\left\{ \int_{Z_1} F(t)\,dt, \int_{Z_2} F(t)\,dt \right\}. \tag{7.5}$$

We have substituted 1 for the diameter of the interval $[0,1]$. Also, $d(Z_1,Z_2) \geq d(S_1,S_2)/D = C/2$. Thus, Theorem 30 applied to the function $F(t)$ contradicts (7.4) and to prove the theorem in general, it suffices to prove it in the one-dimensional case.

In fact, it will be enough to prove (7.5) for the case when each Z_i is a single interval. Suppose we can do this. Then, for each maximal interval (c,d) contained in Z_3, the integral of F over Z_3 is at least C times the smaller of the integrals to its left $[0,c]$ and to its right $[d,1]$ and so one of these intervals is "accounted" for. If all of Z_1 or all of Z_2 is accounted for, then we are done. If not, there is an unaccounted subset U that intersects both Z_1 and Z_2. But then, since Z_1 and Z_2 are separated by at least $d(S_1,S_2)/D$, there is an interval of Z_3 of length at least $d(S_1,S_2)/D$ between $U \cap Z_1$ and $U \cap Z_2$ which can account for more.

We are left with proving (7.5) when each Z_i is an interval. Without the factor of two, this is trivial by the logconcavity of F. To get C as claimed, one can reduce this to the case when $F(t) = e^{ct}$ and verify it for this function [**?**]. The main step is to show that there is a choice of c so that when the current $F(t)$ is replaced by e^{ct}, the LHS of (7.5) does not increase and the RHS does not decrease.