In this chapter, we study polynomial-time sampling algorithms. As a primary motivation for much of the analysis, we consider the *ball walk*.

The Ball Walk

The simplest continuous algorithm for exploring space is Brownian motion with the ODE $dx_t = dW_t$. To turn this into a sampling algorithm, for a convex function f, we saw an extension using the gradient of f, namely

$$dx_t = -\nabla f(x_t) + \sqrt{2} dW_t$$

which can be used to sample according to the density proportional to e^{-f}. In this chapter we will begin with an even simpler method, which does not need access to the gradient or assume differentiability, only an oracle that can evaluate F.

Algorithm 1: `BallWalk`

Input: Step-size δ, number of steps T, starting point x_0 in the support of target density Q.

Repeat T times: at a point x,

1. Pick a random point y in the δ-ball centered at x.
2. Go to y with probability $\min\left\{1, \frac{Q(y)}{Q(x)}\right\}$.

return x.

Exercise 0.1. Show that the distribution with density Q is *stationary* for the ball walk in a connected, compact full-dimensional set, i.e., if the distribution of the current point x has density Q, it remains Q.

Under mild conditions, the distribution of the current point approaches the target density Q. The main question is the rate of convergence, which would allow us to bound the number of steps. Note that each step involves only a function evaluation, to an oracle that outputs the value of a function proportional to the desired density. To bound the rate of convergence (and as a result the uniqueness of the stationary distribution), we first develop some general tools.

0.1 Basics of Markov chains

For more detailed reading, including additional properties, see Section 1 of [**?**].

A Markov chain is defined using a σ-algebra $(K, \mathcal{A})$, where K is the state space and $\mathcal{A}$ is a set of subsets of K that is closed under complements and countable unions. For each element u of K, we have a probability measure P_u on $(K, \mathcal{A})$, i.e., each set $A \in \mathcal{A}$ has a probability $P_u(A)$. Informally, P_u is the distribution obtained upon taking one step from u. The triple $(K, \mathcal{A}, \{P_u : u \in K\})$ along with a starting distribution Q_0 defines a Markov chain, i.e., a sequence of elements of K, $w_0, w_1, \ldots$, where w_0 is chosen from Q_0 and each subsequent w_i is chosen from $P_{w_{i-1}}$. The choice of w_{i+1} depends only on w_i and is independent of $w_0, \ldots, w_{i-1}$.

A distribution Q on $(K, \mathcal{A})$ is called *stationary* if taking one step from it maintains the distribution, i.e., for any $A \in \mathcal{A}$,

$$\int_K P_u(A)\, dQ(u) = Q(A).$$

A distribution Q is *atom-free* if there is no $x \in K$ with $Q(x) > 0$.

Example. For the ball walk in a convex body, the state space K is the convex body, and A is the set of all measurable subsets of K. The next step distribution is

$$P_u(\{u\}) = 1 - \frac{(K \cap (u + \delta B_n))}{(\delta B_n)}$$

and for any measurable subset A,

$$P_u(A) = \frac{(A \cap (u + \delta B_n))}{(\delta B_n)} + 1_{u \in A}(u) P_u(\{u\})$$

The uniform distribution is stationary, i.e., $Q(A) = \frac{(A)}{(K)}$.

The *ergodic flow* of a subset A w.r.t. the distribution Q is defined as

$$\Phi(A) = \int_A P_u(K \setminus A)\, dQ(u).$$

A distribution Q is stationary if and only if $\Phi(A) = \Phi(K \setminus A)$. The existence and uniqueness of the stationary distribution Q for general Markov chains is a subject on its own. One way to ensure uniqueness of a stationary distribution is to use *lazy* Markov chains. In a lazy version of a given Markov chain, at each step, with probability 1/2, we do nothing; with the rest we take a step according to the Markov chain. The next theorem is folklore.

Exercise 0.2. If Q is stationary w.r.t. a lazy, ergodic Markov chain, then it is the unique stationary distribution for that Markov chain.

Informally, the mixing rate of a random walk is the number of steps required to reduce some measure of the distance of the current distribution to the stationary distribution by a constant factor. The following notions will be useful for comparing two distributions P, Q.

1. Total variation distance is $d_{tv}(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$.
2. L_2 or χ^2-distance of P with respect to Q is

$$\chi^2(P, Q) = \int_K \left(\frac{dP(u)}{dQ(u)} - 1\right)^2 dQ(u) = \int_K \left(\frac{dP(u)}{dQ(u)}\right)^2 dQ(u) - 1 = \int_K \frac{dP(u)}{dQ(u)}\, dP(u) - 1.$$

3. KL-divergence:

$$d_{KL}(P, Q) = \int_K \log \frac{dP(u)}{dQ(u)}\, dP(u).$$

4. Warmth: P is said to be *M-warm* w.r.t. Q if $M = \sup_{A \in \mathcal{A}} \frac{P(A)}{Q(A)}$.

Definition 0.3. We say that a distribution P is *absolutely continuous* w.r.t. another distribution Q if for any set A for which $Q(A) = 0$ we also have $P(A) = 0$.

The following relationships hold when P is absolutely continuous w.r.t. Q.

Fact 0.4. *We have*

$$d_{TV}(P, Q)^2 \leq 2 d_{KL}(P, Q) \leq 2\chi^2(P, Q).$$

Convergence via Conductance

Now we introduce an important tool to bound the rate of convergence of Q_t,the distribution after t steps to Q. Assume that Q is the unique stationary distribution. The *conductance* of a subset A is defined as

$$\phi(A) = \frac{\Phi(A)}{\min\{Q(A), Q(K \setminus A)\}}$$

and the conductance of the Markov chain is

$$\phi = \min_A \phi(A) = \min_{0 < Q(A) \leq \frac{1}{2}} \frac{\int_A P_u(K \setminus A)\, dQ(u)}{Q(A)}.$$

The *local* conductance of an element u is $\ell(u) = 1 - P_u(\{u\})$.

For any $0 \leq s < \frac{1}{2}$, the s-conductance of a Markov chain is defined as

$$\phi_s = \min_{A : s < Q(A) \leq \frac{1}{2}} \frac{\phi(A)}{Q(A) - s}.$$

Ideally we would like to show that $d(Q_t, Q)$, the distance between the distribution after t steps and the target Q is monotonically (and rapidly) decreasing. We consider

$$\sup_{A:Q(A)=x} Q_t(A) - Q(A)$$

for each $x \in [0,1]$. To prove inductively that this quantity decreases, Let $\mathcal{G}_x$ be the set of functions defined as

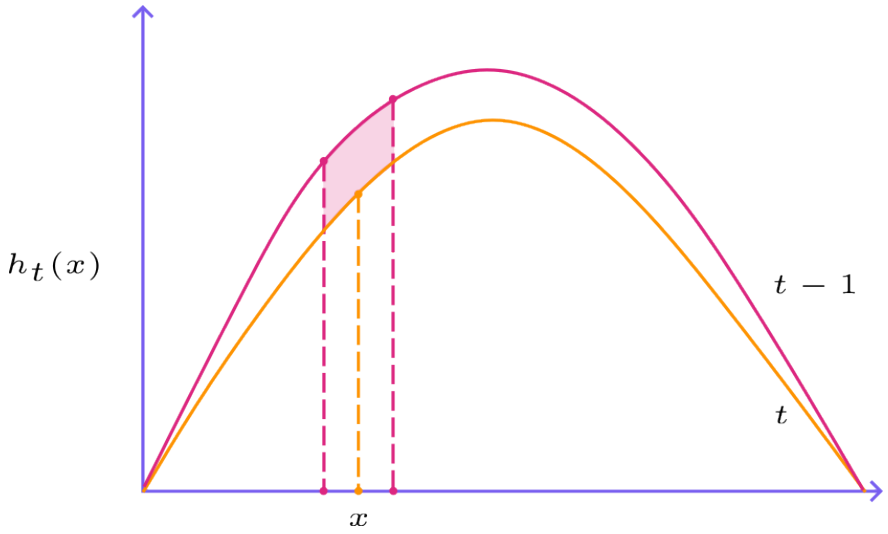
$$\mathcal{G}_x = \left\{ g : K \to [0,1] \ : \ \int_{u \in K} g(u)\ dQ(u) = x \right\}.$$

Using this, define

$$h_t(x) = \sup_{g \in \mathcal{G}_x} \int_{u \in K} g(u)\ (dQ_t(u) - dQ(u)) = \sup_{g \in \mathcal{G}_x} \int_{u \in K} g(u)\ dQ_t(u) - x.$$

This function is concave.

Exercise 0.5. Show that the function h_t is concave, and if Q is atom-free , then $h_t(x) = \sup_{A:Q(A)=x} Q_t(A) - Q(A)$ and the supremum is achieved by some subset.



If the target density Q has atoms, i.e., points of positive probability, then the function g achieving h might have a fractional value at one atom.

Lemma 0.6. *Let Q be atom-free and $t \geq 1$. For any $0 \leq x \leq 1$, let $y = \min\{x, 1-x\}$. 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 $0 \leq x \leq \frac{1}{2}$. We construct two functions, g_1 and g_2, and use these to bound $h_t(x)$. Let A be a subset that achieves $h_t(x)$. Define

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

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

$$\frac{1}{2} \int_{u \in K} g_1(u)\ dQ_{t-1}(u) + \frac{1}{2} \int_{u \in K} g_2(u)\, dQ_{t-1}(u) = \int_{u \in K} P_u(A)\, dQ_{t-1}(u) = Q_t(A).$$

Since the walk is lazy, $P_u(A) \geq \frac{1}{2}$ iff $u \in A$, the range of the functions g_1, g_2 is $[0,1]$. We let

$$x_1 = \int_{u \in K} g_1(u)\ dQ(u) \quad \text{and} \quad x_2 = \int_{u \in K} g_2(u)\ dQ(u),$$

then $g_1 \in \mathcal{G}_{x_1}$ and $g_2 \in \mathcal{G}_{x_2}$. Moreover,

$$\frac{1}{2}(x_1 + x_2) = \frac{1}{2}\int_{u \in K} g_1(u)\ dQ(u) + \frac{1}{2}\int_{u \in K} g_2(u)\ dQ(u) = \int_{u \in K} P_u(A)\ dQ(u) = Q(A) = x.$$

since Q is stationary.

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

Next,

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

Thus we have, $x_1 \leq x(1-2\phi) \leq x \leq x(1+2\phi) \leq x_2$. Since h_{t-1} is concave, the chord from x_1 to x_2 on h_{t-1} lies below the chord from $[x(1-2\phi), x(1+2\phi)]$. Therefore,

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

□

A proof along the same lines implies the following generalization.

Lemma 0.7. *Let Q be atom-free and $0 \leq s \leq 1$. For any $s \leq x \leq 1-s$, let $y = \min\{x-s, 1-x-s\}$. Then for any integer $t > 0$,*

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

These results can be extended to the case when Q has atoms with slightly weaker bounds [**?**].

Theorem 0.8. *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

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

The proof is by induction on t.

Corollary 0.9. *We have*

1. *Let* $M = \sup_A Q_0(A)/Q(A)$. *Then,*

$$d_{TV}(Q_t, Q) \leq \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t.$$

2. *Let* $0 < s \leq \frac{1}{2}$ *and* $H_s = \sup\{|Q_0(A) - Q(A)| : Q(A) \leq s\}$. *Then,*

$$d_{TV}(Q_t, Q) \leq H_s + \frac{H_s}{s}\left(1 - \frac{\phi_s^2}{2}\right)^t.$$

3. *Let* $M = \chi^2(Q_0, Q)$. *Then for any* $\varepsilon > 0$,

$$d_{TV}(Q_t, Q) \leq \varepsilon + \sqrt{\frac{M}{\varepsilon}}\left(1 - \frac{\phi^2}{2}\right)^t.$$

Exercise 0.10. Show that for a lazy, time-reversible Markov chain with stationary distribution Q and conductance ϕ, we have

$$\chi^2(Q_t, Q) \leq \left(1 - \frac{\phi^2}{2}\right)^t \chi^2(Q_0, Q).$$

Convergence via Log-Sobolev

For a warm start, the convergence rate established by conductance is asymptotically optimal in many cases of interest, including the ball walk for convex body. However, when the starting distribution is more focused, e.g., a single point, then there is a significant starting penalty usually a factor of the dimension or larger. One way to avoid this is to observe that the conductance of smaller subsets is in fact even higher and that one does not need to pay this large starting penalty. A classical technique in this regard is the log-Sobolev constant. For a Markov chain with stationary density Q and transition operator P, we can define it as follows.

$$\rho = \inf_{g:\text{smooth},\ \int g(x)^2 dQ(x) = 1} \frac{\int_{x,y\in K} (g(x) - g(y))^2 P(x,y) dQ(x)}{\int g(x)^2 \log g(x)^2 dQ(x)}.$$

This parameter allows us to show convergence of the current distribution to the target in relative entropy. Recall that the relative entropy of a distribution P with respect to a distribution Q is

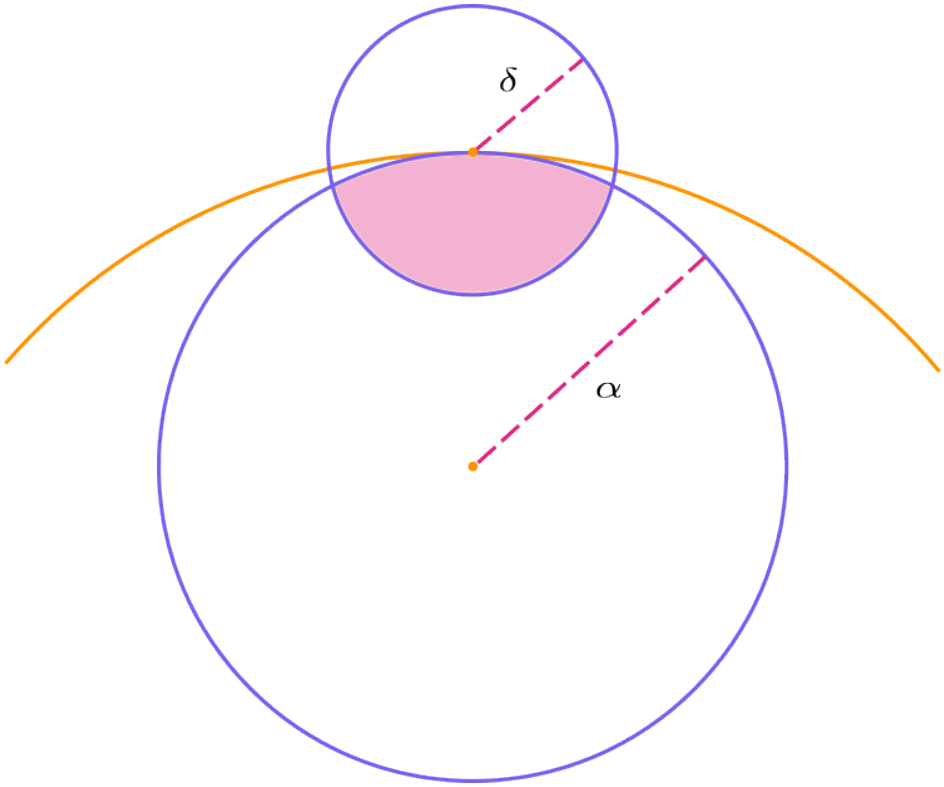
$$H_Q(P) = \int_K P(x) \log \frac{P(x)}{Q(x)} dQ(x).$$

Theorem 0.11. *For a Markov chain with distribution* Q_t *at time* t, *and log-Sobolev parameter* ρ, *we have*

$$H_Q(Q_t) \leq e^{-2\rho t} H_Q(Q_0).$$

0.2 Conductance of the Ball Walk

In the section we bound the conductance of the ball walk when applied to the indicator function of a convex body. At first glance, the ball walk is not an efficient algorithm, even for uniformly sampling a convex body. The reason is simply that the local conductance could be exponentially small (consider a point close to the vertex of a polyhedron). We can get around this in two ways. The first, which is simpler, but less efficient is to "smoothen" the convex body by taking the Minkowski sum with a small Euclidean ball, i.e., replace K with $K + \alpha B^n$.



Exercise 0.12. Let K be a convex body in ^{n} containing the unit ball. Show that (a) $(K+\alpha B^n) \leq (1+\alpha)^n K$ and (b) with $\delta = \alpha/\sqrt{n}$ the local conductance of every point in $K + \alpha B^n$ is at least an absolute constant.

Using the exercise, it suffices to set $\alpha = 1/n$, so that a sample from $K + \alpha B^n$ has a large probability of being in K and then $\delta = 1/n^{3/2}$.

The second approach is to show that the local conductance is in fact large almost everywhere, and if the starting distribution is "warm" then these points can effectively be ignored. This will allow us to make δ much larger, namely $\delta = 1/\sqrt{n}$. Larger step sizes should allows us to prove faster mixing.

To convey the main ideas of the analysis, we focus on the first approach here. The goal is to show that the conductance of any subset is large, i.e., the probability of crossing over in one step is at least proportional to the measure of the set or its complement, whichever is smaller. First, we argue that the one-step distributions of two points will have a significant overlap if the points are sufficient close.

Lemma 0.13 (One-step overlap)**.** *Let $u, v \in K$ s.t. $\ell(u), \ell(v) \geq \ell$ and $u - v \leq \frac{t\delta}{\sqrt{n}}$. Then the one-step distributions from them satisfy $d_{TV}(P_u, P_v) \leq 1 + t - \ell$.*

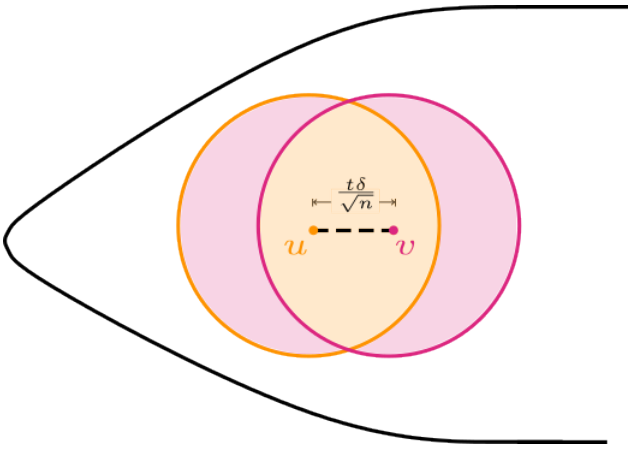


Figure 1: One-step overlap fully contained in K

Proof. First assume that $\ell = 1$. This means that the balls of radius δ centered at u and v are fully contained in K. The TV distance between these distributions is bounded by $(u+\delta B^n \setminus v+\delta B^n) = (v+\delta B^n \setminus u+\delta B^n)$ divided by (δB^n). This is exactly the volume of a band of thickness $u-v$ centered at the center of δB^n (see Fig.). This has relative volume at most t, proving the lemma under the assumption that $\ell = 1$. For the general case, we note that when the balls are intersected with a convex body, the increase in the TV distance is at most the probability that there is no proper move at u (or v), i.e., $1-\ell$. This completes the proof.

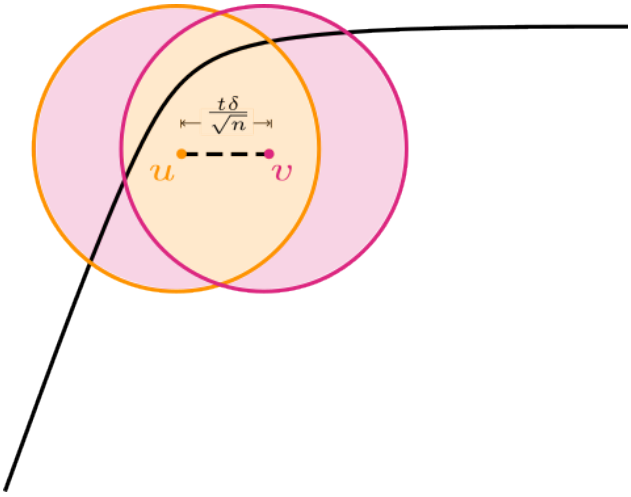


Figure 2: One-step overlap intersected by K

□

Setting $t = \ell/2$, this says that if the total variation distance between the one-step distributions from u, v is greater than $1-\ell/2$, then the distance between them is at least $\frac{\ell\delta}{2\sqrt{n}}$. What this effectively says is that points close to the internal boundary of a subset are likely to cross over to the other side. To complete a proof we would need to show that the internal boundary of any subset is large if the subset (or its complement) is large, a purely geometric property.

Theorem 0.14 (Isoperimetry)**.** *Let S_1, S_2, S_3 be a partition of a convex body K of diameter D. Then,*

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

This can be generalized to any logconcave measure. We will discuss this and other extensions in detail later. But first we bound the conductance.

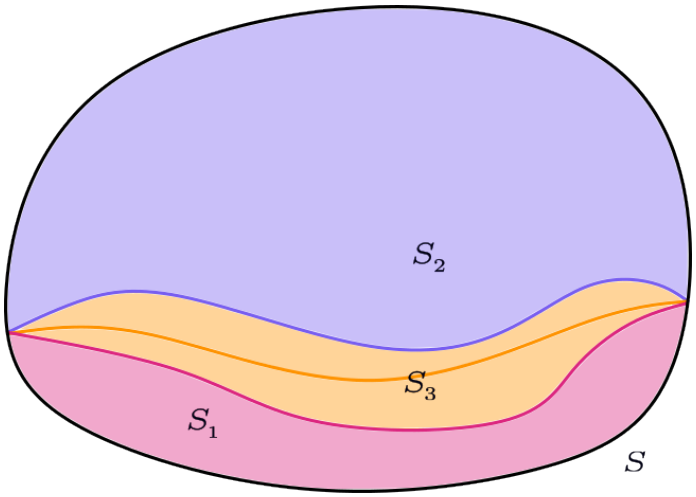


Figure 3: Partitions of convex body

Theorem 0.15. *Let K be a convex body in ^{n} of diameter D containing the unit ball and with every $u \in K$ having $\ell(u) \geq \ell$. Then the conductance of the ball walk on K with step size δ is*

$$\Omega\left(\frac{\ell^2\delta}{\sqrt{n}D}\right).$$

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

$$\int_{S_1} P_x(S_2)\,dx \quad \geq \quad \frac{\ell^2\delta}{16\sqrt{n}D}\min\{(S_1),(S_2)\} \tag{1}$$

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:

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

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

Suppose $(S_1') < (S_1)/2$. Then

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

which proves (1).

So we can assume that $(S_1') \geq (S_1)/2$ and similarly $(S_2') \geq (S_2)/2$. 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 0.13 with $t = \ell/2$, we get that

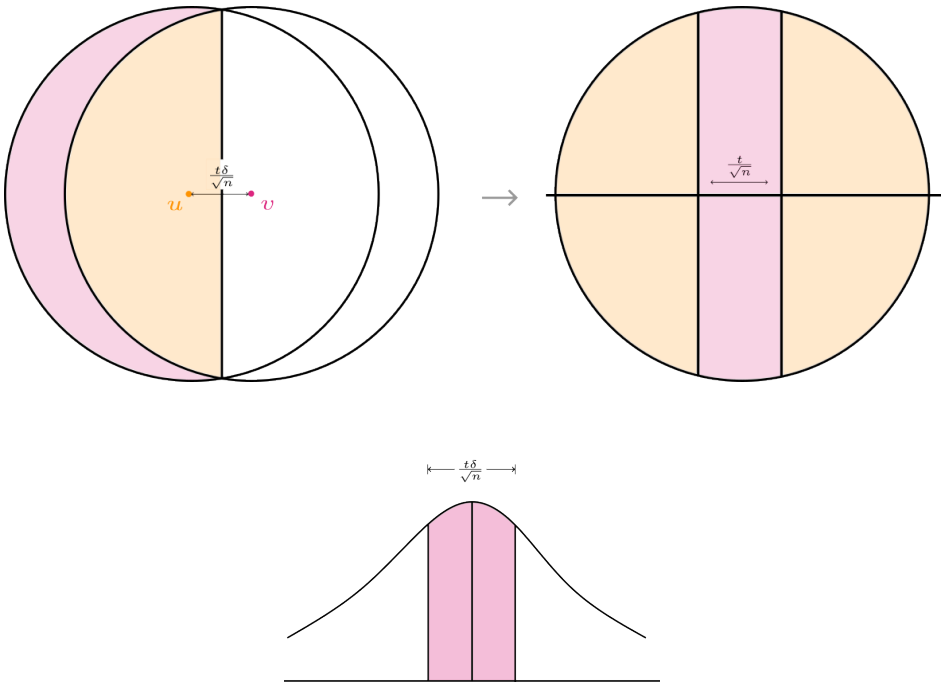
$$|u - v| \geq \frac{\ell\delta}{2\sqrt{n}}.$$

Thus $d(S_1, S_2) \geq \ell\delta/2\sqrt{n}$. Applying Theorem 0.14 to the partition S_1', S_2', S_3', we have

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

We can now prove (1) as follows:

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



□

Corollary 0.16. *The ball walk in a convex body with local conductance at least ℓ everywhere has mixing rate $O(nD^2\delta^2/\ell^4)$.*

Using the construction above of adding a small ball to every point of K gives a lower bound of $\delta = 1/n^{3/2}$ and $\ell = \Omega(1)$ and thus a polynomial bound of $O(n^4D^2)$ on the mixing time. As we will see presently, this can be improved to n^2D^2 by avoiding the blow-up, and analyzing the *average* local conductance. The example of starting near a corner (say of a hypercube) shows that this cannot work in general; however, from a *warm* start, it will suffice to bound the average local conductance rather than the minimum.

Warm Start and s-Conductance

In this section we give a better bound for the conductance of sufficiently large subsets resulting in the following bound on the mixing rate of the ball walk.

Theorem 0.17. *From a warm start, the ball walk in a convex body of diameter D containing a unit ball has a mixing rate of $O(n^2D^2)$ steps.*

This is based on two ideas: (1) most points of a convex body containing a unit ball have large local conductance and we can use $\delta = 1/\sqrt{n}$ instead of $1/n^{3/2}$, (2) the s-conductance is large and hence the walk mixes from a suitably warm start.

Lemma 0.18. *Let K be a convex body containing a unit ball. For the ball walk with δ step size, let $K_\delta = \left\{u \in K : \ell(u) \geq \frac{3}{4}\right\}$. Then K_δ is a convex set and $(K_\delta) \geq (1 - 2\delta\sqrt{n})(K)$.*

Exercise 0.19. Prove the first part of the previous lemma.

The proof of the above lemma will use the following bound on crossing the boundary of K.

Lemma 0.20. *Let L be any measurable subset of the boundary of a convex body K and $S_L = \{(x, y) : x \in K, y \notin K, x - y \leq \delta, [$ Then,, we have*

$$_{2n}(S_L) \leq \frac{\delta}{(n+1)} \frac{(B^{n-1})}{(B^n)} {}_{n-1}(L)(\delta B^n).$$

Proof. It suffices to consider the case when L is infinitesimally small; and then we can assume that the surface of K is locally a hyperplane and compute the measure of S_L explicitly. □

Theorem 0.21. *The s-conductance of the ball walk with $\delta = \frac{s}{4\sqrt{n}}$ step size in a convex body K of diameter D and containing the unit ball satisfies*

$$\phi_s \gtrsim \frac{s}{nD}.$$

This proof is similar to that of Theorem 0.15, with one important extension. Rather than applying the argument to a partition of the original convex body K, we restrict the argument to K_δ,the points in K that have high local conductance. Since this subset takes up most of K, and is convex, we will able to use its isoperimetry to lower bound the conductance.

Proof. As before, we consider the following partition of K. Let $K = S_1 \cup S_2$ be a partition into measurable sets. We will prove that

$$\int_{S_1} P_x(S_2)\, dx \quad \geq \quad \frac{\delta}{C\sqrt{n}D} \min\{(S_1) - \frac{s}{2}, (S_2) - \frac{s}{2}\} \tag{2}$$

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:

$$S_1' = \left\{x \in S_1 : P_x(S_2) < \frac{1}{8}\right\} \text{ and } S_2' = \left\{x \in S_2 : P_x(S_1) < \frac{1}{8}\right\}.$$

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

$$K_\delta = \left\{x \in K : \ell(x) \geq \frac{3}{4}\right\}$$

and define $S_i'' = S_i' \cap K_\delta$. Note that by Lemma 0.18, for $\delta \leq s/(4\sqrt{n})$,

$$(S_i'') \geq (S_i') - s.$$

Suppose $(S_1'') < (S_1 \cap K_\delta)/2$. Then

$$\int_{S_1} P_x(S_2)\, dx \geq \frac{1}{8}(S_1 \cap K_\delta \setminus S_1'') \geq \frac{1}{16}(S_1 \cap K_\delta) \geq \frac{1}{16}\left((S_1) - \frac{s}{2}\right)$$

which proves (2).

So we can assume that $(S_1'') \geq (S_1 \cap K_\delta)/2$ and similarly $(S_2'') \geq (S_2 \cap K_\delta)/2$. Now, for any $u \in S_1''$ and $v \in S_2''$,

$$d_{TV}(P_u, P_v) \geq 1 - P_u(S_2) - P_v(S_1) > 1 - \frac{1}{4}.$$

Applying Lemma 0.13 with $t = 3/8$, we get that

$$|u - v| \geq \frac{3\delta}{8\sqrt{n}}.$$

Thus $d(S_1'', S_2'') \geq 3\delta/(8\sqrt{n})$. Applying Theorem 0.14 to the partition S_1'', S_2'', S_3'' of K_δ, we have

$$\begin{aligned}(S_3'') &\geq \frac{3\delta}{4\sqrt{n}D} \min\{(S_1''), (S_2'')\} \\ &\geq \frac{3\delta}{8\sqrt{n}D} \min\{(S_1) - \frac{s}{2}, (S_2) - \frac{s}{2}\}.\end{aligned}$$

We can now prove (2) as follows:

$$\begin{aligned}\int_{S_1} P_x(S_2)\,dx &= \frac{1}{2}\int_{S_1} P_x(S_2)\,dx + \frac{1}{2}\int_{S_2} P_x(S_1)\,dx \\ &\geq \frac{1}{2}(S_3'') \cdot \frac{1}{8} \\ &\geq \frac{3\delta}{128\sqrt{n}D} \min\{(S_1), (S_2)\}\end{aligned}$$

which implies that the conductance is at least $3s/(512nD)$. □

The mixing rate then follows by applying Theorem 0.8 and Corollary 0.9.

Tightness of the bound

The mixing rate of $O(n^2D^2)$ for the ball walk is in fact the best possible even from a warm start (with the assumption of a unit ball inside the convex body and diameter D). To see this, consider a cylinder whose cross-section is a unit ball and axis is $[0, D]$ along e_1. Suppose the starting distribution is uniform in the part of the cylinder in $[0, D/3]$. Then we claim that to reach the opposite third of the cylinder needs $\Omega(n^2D^2)$ steps with high probability. Each step has length at most δ in a random direction, and this is about $\delta/\sqrt{n}$ along e_1. Viewing this as an unbiased random walk along e_1, the effective diameter is $D/(\delta/\sqrt{n})$ and hence the number of steps to cross an interval of length $D/3$ is $\Omega(nD^2/\delta^2) = \Omega(n^2D^2)$.

Exercise 0.22. Prove the above claim rigorously.

Speedy walk

In the above analysis of the ball walk, the dependence on the error parameter ε, the distance to the target distribution, is polynomial in $1/\varepsilon$ rather than its logarithm. The speedy walk is a way to improve the analysis. In the speedy walk, at a point x, we sample the next step uniformly from the intersection of $(x + \delta B^n) \cap K$. The resulting Markov chain is the subsequence of *proper* steps of the ball walk.

Exercise 0.23. Show that the stationary density of the speedy walk in a convex body is proportional to the local conductance.

Theorem 0.24. *The conductance of the speedy walk is* $\Omega(1/nD)$.

To analyze the ball walk, we then need to show that the number of "wasted" steps is not too many. This follows from the assumption of a warm start and Lemma 0.18.

0.3 Generating a warm start

To get the mixing rate of $O(n^2D^2)$, we need a warm start, i.e., a distribution whose density at any point is within $O(1)$ of the target density.

Algorithm 2: `WarmStart`

Input: membership oracle for K s.t. $B^n \subseteq K \subseteq DB^n$.
Let x be a random point in B^n. Define $K_i = 2^{i/n}B^n \cap K$.
for $i = 1, \cdots, n \log D$ **do**
  1. Use ball walk from x to generate random point y in K_i.
  2. Set $x = y$.

end
return x.

Since $K_{i+1} \subseteq 2^{1/n}K_i$, we have $(K_{i+1}) \leq 2(K_i)$ and hence a 2-warm start is maintained. Once we have a random point from K, subsequent random points can be generated by simply continuing the ball walk; thus the cost of the warm start is only for the first sample.

0.4 Isotropic Transformation

The complexity of sampling with the ball walk is polynomial in n, D and $\log(1/\varepsilon)$ to get within ε of the target density. This is not a polynomial algorithm since the dependence is on D and not $\log D$. To get a polynomial algorithm, we need one more ingredient.

We say that a distribution Q is *isotropic* if $_Q(x) = 0$ and $_Q(xx^T) = I$, i.e., the mean is zero and the covariance matrix (exists and) is the identity. We say that the distribution is C-isotropic if the eigenvalues of its covariance matrix are in $[\frac{1}{C}, C]$. An affine transformation is said to be an *isotropic tranformation* if the resulting distribution is isotropic.

Any distribution with bounded second moments has an isotropic transformation. It is clear that satisfying the first condition is merely a translation, so assume the mean is zero. For the second, suppose the covariance matrix is $_Q(xx^T) = A$. Then consider $y = A^{-1/2}x$. It is easy to see that

$$(yy^T) = A^{-1/2}\left(xx^T\right)A^{-1/2} = I.$$

For convex bodies, isotropic position comes with a strong guarantee.

Theorem 0.25. *For a convex body in isotropic position (i.e., the uniform distribution over the body is isotropic), we have*

$$\sqrt{\frac{n+2}{n}}B^n \subseteq K \subseteq \sqrt{n(n+2)}B^n.$$

Thus the effective diameter is $O(n)$. If we could place a convex body in isotropic position before sampling, we would have a poly(n) algorithm. In fact, it is even better than this as most points are within distance $O(\sqrt{n})$ of the center of gravity. We quote a theorem due to Paouris.

Theorem 0.26. *For an isotropic logconcave density p in ^{n} and any $t \geq 1$,*

$$\Pr_p\left(x \geq ct\sqrt{n}\right) \leq e^{-t\sqrt{n}}.$$

How to compute an isotropic transformation? This is easy, from the definition, all we need is to estimate its covariance matrix, which can be done from random samples. Thus, if we could sample K, we can compute an isotropic transformation for it. This appears cyclic – we need isotropy for efficient sampling and efficient sampling for isotropy. The solution is simply to bootstrap them.

Algorithm 3: `IsotropicTransform`

Input: membership oracle for K s.t. $B^n \subseteq K \subseteq DB^n$.
Let x be a random point in B^n, $A = I$ and $K_i = 2^{i/n}B^n \cap K$.
for $i = 1, \cdots, n \log D$ **do**
1. Use the ball walk from x to generate N random points $x_1 \dots x_N$ in AK_i.
2. Compute $C = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T$ and set $A = C^{-1/2}A$.
3. Set $x = x_N$.

end
return x.

We will choose N large enough so that after the transformation K_i is 2-isotropic and therefore K_{i+1} is 6-isotropic. We can bound N as follows.

Exercise 0.27. Show that if K is isotropic, then with $N = O(n^2)$, the matrix $A = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T$ for N random samples from K satisfies $A - I \leq 0.5$.

A tight bound on the sample complexity was established by [?] (see also [?, ?, ?]).

Theorem 0.28. *For an isotropic logconcave distribution Q in ^{n}, the covariance $N = O(n)$ random samples satisfies $A - I \leq 0.5$.*

Thus the overall algorithm needs $O(n \log D)$ phases, with $O(n)$ samples in each phase from a near-isotropic distribution, and thus poly(n) steps per sample.

0.5 Isoperimetry via localization

Theorem 0.14 was refined by KLS [?] as follows (we state it here for logconcave densities).

Theorem 0.29. *For any partition S_1, S_2, S_3 of ^{n}, and any logconcave measure μ in ^{n},*

$$\mu(S_3) \geq \frac{\ln 2}{_\mu(x - \bar{x})} \min\{\mu(S_1), \mu(S_2)\}.$$

Thus, for a (near-)isotropic distribution, the diameter can be replaced by $O(\sqrt{n})$ and this gives a bound of $O(n^3)$ from a warm start. One way to summarize the analysis so far is that the complexity of sampling a convex body (and in fact a logconcave density) from a warm start is $O^*(n^2/\psi^2)$ where ψ is the isoperimetric ratio of the convex body. In other words, the expansion of the Markov chain reduces to the expansion of the target logconcave density. It then becomes a natural question to find the best possible estimate for the isoperimetric ratio. KLS also provided a conjecture for this.

Conjecture 0.30 (KLS Hyperplane Conjecture.)**.** *The isoperimetric ratio of any isotropic logconcave density in ^{n} is $\Omega(1)$.*

The bound of the conjecture holds for all halfspace induced subsets. So the conjecture says that the worst isoperimetry is achieved up to a constant factor by a halfspace (this version does not need isotropic position). Here we discuss a powerful technique for proving such inequalities.

Classical proofs of isoperimetry for special distributions are based on different types of symmetrization that effectively identify the extremal subsets. Bounding the Cheeger constant for general convex bodies and logconcave densities is more complicated since the extremal sets can be nonlinear and hard to describe precisely, due to the trade-off between minimizing the boundary measure of a subset and utilizing as much of the "external" boundary as possible. The main technique to prove bounds in the general setting has been *localization*, a method to reduce inequalities in high dimension to inequalities in one dimension. We now describe this technique with a few applications.

0.5.1 Localization

We will sketch a proof of the following theorem to illustrate the use of localization. This theorem was also proved by Karzanov and Khachiyan [?] using a different, more direct approach.

Theorem 0.31 ([?, ?, ?]). *Let f be a logconcave function whose support has diameter D and let π_f be the induced measure. Then for any partition of ^{n} into measurable sets S_1, S_2, S_3,*

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

Before discussing the proof, we note that there is a variant of this result in the Riemannian setting.

Theorem 0.32 ([?]). *If $K \subset (M, g)$ is a locally convex bounded domain with smooth boundary, diameter D and $Ric_g \geq 0$, then the Poincaré constant is at least $\frac{\pi^2}{4D^2}$, i.e., for any g with $\int g = 0$, we have that*

$$\int |\nabla g(x)|^2 \, dx \geq \frac{\pi^2}{4D^2} \int g(x)^2 dx.$$

For the case of convex bodies in , this result is equivalent to Theorem 0.31 up to a constant. One benefit of localization is that it does not require a carefully crafted potential. Localization has been extended to Riemannian setting [?]. The origins of this method were in a paper by Payne and Weinberger [?].

We begin the proof of Theorem 0.31. For a proof by contradiction, let us assume the converse of its conclusion, i.e., for some partition S_1, S_2, S_3 of ^{n} and logconcave density f, assume that

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

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

$$\int_n g(x)\,dx > 0 \quad \text{and} \quad \int_n h(x)\,dx > 0 \tag{3}$$

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 ^{n}. The next lemma will help us analyze them.

Lemma 0.33 (Localization Lemma [?]). *Let $g, h :^n \to$ be lower semi-continuous integrable functions such that*

$$\int_n g(x)\,dx > 0 \quad \textit{and} \quad \int_n h(x)\,dx > 0.$$

Then there exist two points $a, b \in^n$ and an affine function $\ell : [0,1] \to_+$ 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 and one may think of $\ell(t)^{n-1}$ as proportional to the cross-sectional area of an infinitesimal cone. 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.

Proof outline. The main idea is the following. Let H be any halfspace such that

$$\int_H g(x)\,dx = \frac{1}{2} \int_n g(x)\,dx.$$

Let us call this a bisecting halfspace. Now either

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

Thus, either H or its complementary halfspace will have positive integrals for both g and h, reducing the domain of the integrals from ^{n} to a halfspace. If we could repeat this, we might hope to reduce the dimensionality of the domain. For any $(n-2)$-dimensional affine subspace L, there is a bisecting halfspace containing L in its bounding hyperplane. To see this, let H be a halfspace containing L in its boundary. Rotating H about L we get a family of halfspaces with the same property. This family includes H', the complementary halfspace of H. The function $\int_H g - \int_{^n\setminus H} g$ switches sign from H to H'. Since this is a continuous family, there must be a halfspace for which the function is zero.

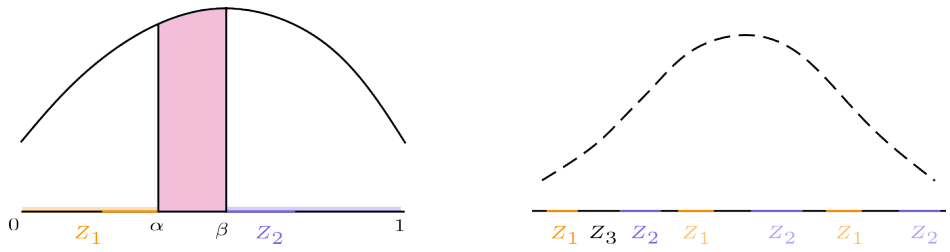
If we take all $(n-2)$-dimensional affine subspaces defined by $\{x \in:\ x_i = r_1,\ x_j = r_2\}$ where r_1, r_2 are rational, then the intersection of all the corresponding bisecting halfspaces is a line or a point (by choosing only rational values for x_i, we are considering a countable intersection). To see why it is a line or a point, assume we are left with a two or higher dimensional set. Since the intersection is convex, 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.

Thus the limit of this bisection process is a function supported on an interval (which could be a single point), and since the function itself is a limit of convex sets (intersections of halfspaces) containing this interval, it is a limit of a sequence of concave functions and is itself concave, with positive integrals. Simplifying further from concave to linear takes quite a bit of work. For the full proof, we refer the reader to [?]. □

Going back to the proof sketch of Theorem 0.31, we can apply the localization lemma to get an interval $[a, b]$ and an affine function ℓ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a + tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a + tb)\,dt > 0. \tag{4}$$

The functions g, h as we have defined them are not lower semi-continuous. However, this can be addressed 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, these modifications do not affect the conclusion.



Let us partition $[0, 1]$ into Z_1, Z_2, Z_3 as follows:

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

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

$$\int_{Z_3} F(t)\,dt < C \min\left\{\int_{Z_1} F(t)\,dt, \int_{Z_2} F(t)\,dt\right\}. \tag{5}$$

Now consider what Theorem 0.31 asserts for the function $F(t)$ over the interval $[0,1]$ and the partition Z_1, Z_2, Z_3:

$$\int_{Z_3} F(t)\,dt \geq 2d(Z_1, Z_2) \min\left\{\int_{Z_1} F(t)\,dt, \int_{Z_2} F(t)\,dt\right\}. \tag{6}$$

We have substituted 1 for the diameter of the interval $[0,1]$. Also, $2d(Z_1, Z_2) \geq 2d(S_1, S_2)/D = C$. Thus, Theorem 0.31 applied to the function $F(t)$ contradicts (5) and to prove the theorem in general, and it suffices to prove it in the one-dimensional case. A combinatorial argument reduces this to the case when each Z_i is a single interval. Proving the resulting inequality up to a factor of 2 is a simple exercise and uses only the unimodality of F. The improvement to the tight bound requires one-dimensional logconcavity. This completes the proof of Theorem 0.31.

The localization lemma has been used to prove a variety of isoperimetric inequalities. The next theorem is a refinement of Theorem 0.31, replacing the diameter by the square-root of the expected squared distance of a random point from the mean. For an isotropic distribution this is an improvement from n to $\sqrt{n}$. This theorem was proved by Kannan-Lovász-Simonovits in the same paper in which they proposed the KLS conjecture.

Theorem 0.34 ([?]). *For any logconcave density p in ^{n} with covariance matrix A, the KLS constant satisfies*

$$\psi_p \gtrsim \frac{1}{\sqrt{(A)}}.$$

The next theorem shows that the KLS conjecture is true for an important family of distributions. The proof is again by localization [?], and the one-dimensional inequality obtained is a Brascamp-Lieb Theorem. We note that the same theorem can be obtained by other means [?].

Theorem 0.35. *Let $h(x) = f(x)e^{-\frac{1}{2}x^\top Bx} / \int f(y)e^{-\frac{1}{2}y^\top By}dy$ where $f :^n \to_+$ is an integrable logconcave function and B is positive definite. Then h is logconcave and for any measurable subset S of ^{n},*

$$\frac{h(\partial S)}{\min\{h(S), h(^n \setminus S)\}} \gtrsim \frac{1}{B^{-1\,\frac{1}{2}}}.$$

In other words, the expansion of h is $\Omega(B^{-1^{-\frac{1}{2}}})$.

The analysis of the Gaussian Cooling algorithm for volume computation [?] uses localization.

Next we mention an application to the anti-concentration of polynomials. This is a corollary of a more general result by Carbery and Wright.

Theorem 0.36 ([?]). *Let q be a degree d polynomial in ^{n}. Then for a convex body $K \subset^n$ of volume 1, any $\epsilon > 0$, and x drawn uniform from K,*

$$\Pr_{x\sim K}\left(|q(x)| \leq \epsilon \max_K |q(x)|\right) \lesssim \epsilon^{\frac{1}{d}} d$$

■ 0.5.2 Alternative formulations of Localization

Here we give some alternative equivalent version of Lemma 0.33 which are sometimes more convenient in applications.

The first is in terms of single inequality on products of integrals.

Lemma 0.37 (Localization: Product form]).

The next is in terms of exponential needles, i.e., we replace the linear functions obtained in the one-dimensional case in Lemma 0.33 with truncated exponentials.

Lemma 0.38 (Localization: Exponential Needles]). *Let $f_1, f_2, f_3, f_4 : [a,b] \to_+$ be nonnegative, continuous integrable functions. Then the following are equivalent:*

1. For every logconcave function $F :\to_+$, we have

$$\int_a^b F(x)f_1(x)\,dx \int_a^b F(x)f_2(x)\,dx \le \int_a^b F(x)f_3(x)\,dx \int_a^b F(x)f_4(x)\,dx.$$

2. For every subinterval $[a',b'] \subseteq [a,b] \in$ and every $\gamma \in$, we have

$$\int_{a'}^{b'} e^{\gamma t}f_1(t)\,dt \int_{a'}^{b'} e^{\gamma t}f_2(t)\,dt \le \int_{a'}^{b'} e^{\gamma t}f_3(t)\,dt \int_{a'}^{b'} e^{\gamma t}f_4(t)\,dt.$$

We conclude this section with a nice interpretation of the localization lemma by Fradelizi and Guedon. They also give a version that extends localization to multiple inequalities.

Theorem 0.39 (Reformulated Localization Lemma [**?**]). *Let K be a compact convex set in and f be an upper semi-continuous function. Let P_f be the set of logconcave distributions μ supported by K satisfying $\int f d\mu \ge 0$. The set of extreme points of $convP_f$ is exactly:*

1. *the Dirac measure at points x such that $f(x) \ge 0$, or*
2. *the distributions v satisfies*
   (a) *density function is of the form e^ℓ with linear ℓ,*
   (b) *support equals to a segment $[a,b] \subset K$,*
   (c) *$\int f dv = 0$,*
   (d) *$\int_a^x f dv > 0$ for $x \in (a,b)$ or $\int_x^b f dv > 0$ for $x \in (a,b)$.*

Since we know the maximizer of any convex function is at extreme points, this shows that one can optimize $\max_{\mu \in P_f} \Phi(\mu)$ for any convex Φ by only checking Dirac measures and log-affine functions.