

The Poincaré Constant of a Random Walk in High-Dimensional Convex Bodies

(MASTERS THESIS)

Ivona Bezáková

Thesis Advisor: Prof. Eric Vigoda

December, 2002

Abstract

Estimating the volume of a convex body is an important algorithmic problem. High dimensions pose an especially difficult task for the algorithm designer. To be considered efficient, the running time must be polynomial in the dimension. The first provably efficient approximation algorithm was presented by Dyer, Frieze and Kannan in 1989, and steadily improved by various authors over the next decade. Fundamental to these works is the analysis of random walks for generating a random point from a convex body.

Kannan, Lovász, and Simonovits recently analyzed a natural random walk known as the ball walk. By bounding the so-called conductance, they obtain (via a Cheeger-type inequality) a bound on the Poincaré constant of the random walk. Their bound proves the walk converges to a random point in time $O^*(n^3 D^2)$, where n is the dimension and D is the diameter of the body. We survey and present a self-contained view of their work. In addition, following an outline proposed by Jerrum, we slightly modify the KLS analysis to bound the Poincaré constant without using a Cheeger-type inequality.

1 Introduction

Sampling points uniformly at random from high-dimensional convex bodies is a natural problem with many applications such as approximating the volume of a convex body.

Throughout this text $K \subseteq \mathbb{R}^n$ denotes an n -dimensional convex body of diameter D . The body K as well as a precision ε is given to the sampling algorithm and it produces a point, or a set of (almost) independent points, selected (almost) uniformly at random from K . The proximity to the uniform distribution is captured by the precision ε and the algorithm is required to run in (expected) polynomial time in n and $1/\varepsilon$. (In fact, the algorithm presented here runs in time polynomial in $\log(1/\varepsilon)$.) To avoid implementation issues the body K is given by a *membership oracle* where for every point we can ask if the point is in the body. Time is measured in the number of oracle calls.

A straightforward Monte Carlo approach wraps the convex body into a rectangular box and samples points from this box until it finds one in the body. The output is a random point from the body, but the expected time needed to generate a point grows exponentially with dimension n .

The first polynomial time randomized algorithm was obtained by Dyer, Frieze, and Kannan in 1989 [5] and it was based on a random walk on a fine grid embedded in K . Their result, an $O^*(n^{23})$ running time for the volume application, was improved several times over the next eight years by various combinations of these authors (listed in alphabetical order): Applegate, Dyer, Frieze, Kannan, Lovász, Simonovits, see [13], [2], [12], [4], [14], [15], and [11]. The latest algorithm by Kannan, Lovász, and Simonovits runs in time $O^*(n^5)$. (The O^* -notation, or the “soft- O ” notation hides logarithmic factors and other factors independent of n .) It is based, as several previous algorithms, on a ball walk approach. The new analysis of the ball walk Markov chain grants $O^*(n^3 D^2)$ running time for sampling. The authors are able to transform the body into an (almost) isotropic position to guarantee low diameter and obtain $O^*(n^5)$ time bound.

Basic idea of all the algorithms is to design a Markov chain with proper limiting distribution, ideally the uniform distribution, and a good convergence rate.

The ball walk Markov chain is a random walk on the convex body where for fixed $\delta > 0$ in one step we can move to a point distant at most δ from the current point. There are two important variants of the ball walk – the speedy walk and the Metropolis walk. In the speedy walk we choose the next point uniformly from all points in the body distant at most δ . More precisely, if x is the current point and $B(x, \delta)$ is the set of all points at most δ away from x (a ball of radius δ centered at x), then the next point is chosen uniformly at random from $B(x, \delta) \cap K$. If δ is larger than the diameter, we could sample points from the body in single step. This points to implementation problems with the speedy walk. The most natural practical implementation chooses a point uniformly at random from $B(x, \delta)$ and then tests whether the chosen point is in K . If not, it stays at the current point, otherwise it moves to the new point. This random walk is named the Metropolis version of the ball walk, shortly the Metropolis walk. The goal is to bound the expected number of steps until we reach a random point from the body. The key argument in the analysis is to bound the number of the speedy and Metropolis steps separately. There is a balance between the size of δ and the number of the speedy vs. Metropolis steps. The larger δ , the fewer speedy steps we need. On the other hand, the smaller δ , the fewer Metropolis steps are required. The optimal δ (with respect to the currently best analysis) is of order $1/\sqrt{n}$. The limiting distribution of neither of the ball walks is uniform but they can be used as a stepping stone to obtaining the uniform distribution.

This thesis presents a self-contained presentation of the $O^*(n^3 D^2)$ result from [11]. The original paper bounds the conductance of the Markov chain, whereas in this work we bound the Poincaré

constant of the chain, formalizing an argument outlined by Jerrum in [8].

Conductance is a combinatorial tool for bounding the convergence rate. It corresponds to the probability of escaping from a worst-case set in one step of the chain. From a functional analysis perspective, conductance is a special case of the Poincaré constant, see Section 2 for more details.

The Poincaré constant captures the speed of decaying of the variance (with respect to the chain’s limiting distribution) of the worst-case random variable on the sample space. The spectral gap of a Markov chain defined on a finite sample space is the difference between the first and second largest (in absolute value) eigenvalues of the chain’s transition matrix. For a reversible Markov chain with finite sample space, the Poincaré constant equals the spectral gap. Thus, the Poincaré constant may be viewed as a convenient generalization of the spectral gap for non-reversible Markov chains and infinite sample spaces. It is a classical result that the spectral gap (or the Poincaré constant) provides a bound on the convergence rate of the chain to its limiting distribution. Exact definitions and statements of the theorems are presented in Section 2.

A Cheeger-type inequality by Jerrum and Sinclair [9] relates the two quantities: Obtaining a bound on the conductance provides a bound on the Poincaré constant (and thus a bound on the convergence rate). In the meantime, Mihail [16] obtained the convergence rate from conductance without the use of the Poincaré constant. However, proving a Poincaré inequality directly avoids Cheeger-type analysis and adds little (if any) difficulty to the proof. In addition, there are several examples where such direct proofs lead to improved convergence results (via the so-called log-Sobolev inequality), see e.g. Frieze and Kannan [7], Jerrum and Son [10].

In Section 2 we review basic definitions and theorems from the theory of Markov chains. In Section 3 we formally define both ball walks and analyze their limiting distributions. The heart of this thesis, Section 4, is devoted to the analysis of the Poincaré inequality of the speedy walk. We conclude with Section 5 where we estimate the number of Metropolis steps and present a trick used for obtaining a point uniform at random from a point sampled according to the limiting distribution of the Metropolis walk.

2 Preliminaries – (Continuous) Markov Chains

We start with definition and properties of a discrete-time Markov chain on (possibly infinite) sample space.

Definition 1 A Markov chain $\mathcal{M} = (\Omega, P)$ on sample space Ω is defined by its transition function $P : \Omega^2 \rightarrow [0, 1]$. For any state $x \in \Omega$, $P(x, \cdot)$ is the probability distribution on Ω determining the likelihood of the next state after taking one step from x . Distribution π is said to be stationary if $\pi P = \pi$.

Markov chain is reversible, if there exists a distribution π such that $\pi(x)P(x, y) = \pi(y)P(y, x)$ for all $x, y \in \Omega$.

When using Markov chains for sampling, we want to design a chain with unique stationary distribution compatible with the sampling’s distribution.

One can easily verify that if a Markov chain is reversible, this π is the stationary distribution. Another useful observation is that if P is symmetric (i.e. $P(x, y) = P(y, x)$ for all x, y), then this Markov chain is reversible with uniform stationary distribution.

Intuitively, mixing time of a Markov chain is its convergence rate, i.e. number of steps required to get sufficiently close to the stationary distribution. Before we state the formal definition, we need to be able to measure distance of two distributions.

Definition 2 Total variation distance of two distributions π and μ is defined as

$$d_{tv}(\pi, \mu) := \sup_{A \subseteq \Omega} |\pi(A) - \mu(A)|$$

Definition 3 Let $\varepsilon > 0$. We say that a Markov Chain with stationary distribution μ has mixing time $t = t(\varepsilon)$ if for a given initial distribution μ_0

$$d_{tv}(\mu_t, \mu) \leq \varepsilon,$$

where the initial state is drawn according to μ_0 and μ_t is the probability distribution after t steps of the chain.

Surprisingly, the total variation distance is not always the best choice of measure. As we will see shortly, Theorem 6 can be simply stated in terms of the so-called L_2 -distance but no equivalent form in terms of total variation distance is available.

Definition 4 The L_2 -distance of two probability distributions π and μ is defined as $\|\pi - \mu\|_2 := \sqrt{\int_{\Omega} \frac{(\pi(x) - \mu(x))^2}{\mu(x)} dx}$.

Simple computation shows that $\|\pi - \mu\|_2 = \sqrt{\int_{\Omega} \frac{\pi(x)^2}{\mu(x)} dx - 1}$. We will use this equivalent definition of the L_2 -distance several times.

Notice that the L_2 -distance is not a distance in proper mathematical sense since it is not symmetric. However, it has some nice properties, the most important one for our application is summarized in Theorem 6.

Using the continuous variant of the Cauchy-Schwarz inequality we prove that the L_2 -distance is at least twice the total variation distance. It is easy to see that $d_{tv}(\pi, \mu) = \frac{1}{2} \int_{\Omega} |\pi(x) - \mu(x)| dx$.

$$\begin{aligned} \|\pi - \mu\|_2 &= \sqrt{\int_{\Omega} \frac{(\pi(x) - \mu(x))^2}{\mu(x)} dx} \\ &= \sqrt{\int_{\Omega} \frac{(\pi(x) - \mu(x))^2}{\mu(x)} dx} \sqrt{\int_{\Omega} \pi(x) dx} \\ &\geq \int_{\Omega} \frac{|\pi(x) - \mu(x)|}{\sqrt{\pi(x)}} \sqrt{\pi(x)} dx \\ &= 2d_{tv}(\pi, \mu) \end{aligned} \tag{1}$$

Definition 5 Let (Ω, P) be a Markov chain with stationary distribution μ . Given a measurable function $f : \Omega \rightarrow \mathbb{R}$ we define the expected value of f as $E_{\mu}(f) := \int_{\Omega} \mu(x)f(x) dx$, the variance of f as

$$\text{Var}_{\mu}(f) := \int_{\Omega} \mu(x)(f(x) - E_{\mu}(f))^2 dx,$$

and the Dirichlet form of f as

$$\mathcal{E}_\mu(f, f) := \int_\Omega \mu(x)h(x) dx, \quad \text{where} \quad h(x) := \frac{1}{2} \int_\Omega P(x, y)(f(x) - f(y))^2 dy.$$

Intuitively, the Dirichlet form is a variation within one step (local variation) whereas variance is the global variation. In this paper we work with discrete-time Markov chains, i.e. every step takes one time unit. One can also define continuous-time Markov chains, where the Dirichlet form simply corresponds to the derivative of the variance with respect to time.

The *lazy variant* of a Markov chain (Ω, P) is the same Markov chain except that before making a step it tosses a coin and with probability $\frac{1}{2}$ it stays in the same state. In other words, the lazy Markov chain (Ω, P_{ZZ}) is defined as $P_{ZZ} := \frac{1}{2}(I + P)$, where $I(x, x) = 1$ and $I(x, y) = 0$ for every $x \neq y$. It is straightforward to see that the important properties of the chain such as stationary distribution and reversibility do not change by making it lazy. Intuitively, the mixing time of a lazy chain doubles compared to the mixing time of the original chain.

For a moment let us consider a Markov chain defined on a finite sample space. By the Frobenius-Peron Theorem the absolute value of all its eigenvalues is upper-bounded by 1. Then the eigenvalues of the lazy variant are guaranteed to be non-negative and there exists a unique eigenvector with eigenvalue 1 corresponding to the stationary distribution of the chain. Thus, it is easier to bound the spectral gap of the lazy variant. The following theorem applies a similar argument to the Poincaré constant.

Theorem 6 *Let λ be a constant and $\mathcal{M} = (\Omega, P_{ZZ})$ be the lazy variant of a reversible Markov chain with stationary distribution μ satisfying the Poincaré inequality: For any measurable function $f : \Omega \rightarrow \mathbb{R}$,*

$$\mathcal{E}_\mu(f, f) \geq \lambda \text{Var}_\mu(f)$$

Then

$$\|\mu_t - \mu\|_2 \leq \left(1 - \frac{\lambda}{2}\right)^t \|\mu_0 - \mu\|_2$$

Definition 7 *Supremum over all λ satisfying the Poincaré inequality is denoted $\lambda_{\mathcal{M}}$ and called the Poincaré constant of the chain \mathcal{M} . Formally,*

$$\lambda_{\mathcal{M}} := \inf_{f: \Omega \rightarrow \mathbb{R}} \frac{\mathcal{E}_\mu(f, f)}{\text{Var}_\mu(f)}.$$

If we view the Dirichlet form as the derivative of the variance, then the Poincaré constant is a measure of decay of the variance.

If the sample space of the Markov chain is finite, the Poincaré constant equals the spectral gap of the chain. For a proof via variation characterization of the second largest eigenvalue, see Aldous and Fill [1, Chapter 3, Sections 4 and 6].

Proof of Theorem 6:

We will prove that $\|\mu_{t+1} - \mu\|_2 \leq \left(1 - \frac{\lambda}{2}\right) \|\mu_t - \mu\|_2$ for $t \geq 0$. In fact, following an argument by Mihail we will prove something stronger: For any measurable function f , $\text{Var}_\mu(P_{ZZ}f) \leq \text{Var}_\mu(f) - \frac{1}{2}\mathcal{E}_\mu(f, f)$, where $P_{ZZ}f(x) = \int_\Omega P_{ZZ}(x, y)f(y) dy$. This statement trivially implies

$$\text{Var}_\mu(P_{ZZ}f) \leq \left(1 - \frac{\lambda}{2}\right) \text{Var}_\mu(f) \tag{2}$$

and we will show shortly how to get from here to the L_2 -distance of μ_t and μ .

Notice that the variance nor Dirichlet form changes if we shift f by a constant. So we may as well assume $E_\mu(f) = 0$. We want to express $\text{Var}_\mu(f) - \text{Var}_\mu(P_{ZZ}f)$. Equivalent definition of variance brings us closer to the Dirichlet form (recall that μ is the stationary distribution of both (Ω, P) and (Ω, P_{ZZ}) and thus $\mu(y) = \int_\Omega \mu(x)P(x, y) dx$):

$$\begin{aligned}\text{Var}_\mu(f) &= \frac{1}{2} \int_\Omega \mu(x)f(x)^2 dx + \frac{1}{2} \int_\Omega \mu(y)f(y)^2 dy \\ &= \frac{1}{2} \int_\Omega \mu(x)f(x)^2 \int_\Omega P(x, y) dy dx + \frac{1}{2} \int_\Omega \int_\Omega \mu(x)P(x, y) dx f(y)^2 dy \\ &= \frac{1}{2} \int_\Omega \int_\Omega \mu(x)P(x, y)(f(x)^2 + f(y)^2) dx dy\end{aligned}$$

In order to express $\text{Var}_\mu(P_{ZZ}f)$ in a form similar to the Dirichlet form, we need another expression for $P_{ZZ}f(x)$:

$$\begin{aligned}P_{ZZ}f(x) &= \int_\Omega \left(\frac{1}{2}I(x, y) + \frac{1}{2}P(x, y) \right) f(y) dy \\ &= \frac{1}{2}f(x) + \frac{1}{2} \int_\Omega P(x, y)f(y) dy \\ &= \frac{1}{2} \int_\Omega P(x, y)(f(x) + f(y)) dy\end{aligned}$$

Since $E_\mu(f) = 0$ and the stationary distribution of a Markov chain and its corresponding lazy chain is the same, for $E_\mu(P_{ZZ}f)$ we get

$$\begin{aligned}E_\mu(P_{ZZ}f) &= \int_\Omega \mu(x)P_{ZZ}f(x) dx = \int_\Omega \mu(x) \int_\Omega P_{ZZ}(x, y)f(y) dy dx = \\ &= \int_\Omega f(y) \int_\Omega \mu(x)P_{ZZ}(x, y) dx dy = \int_\Omega f(y)\mu(y) dy = E_\mu(f) = 0\end{aligned}$$

For $\text{Var}_\mu(P_{ZZ}f)$ this means (using Cauchy-Schwarz inequality in the middle of the derivation)

$$\begin{aligned}\text{Var}_\mu(P_{ZZ}f) &= \int_\Omega \mu(x) \left[\frac{1}{2} \int_\Omega P(x, y)(f(x) + f(y)) dy \right]^2 dx \\ &= \frac{1}{4} \int_\Omega \mu(x) \left[\sqrt{P(x, y)}(f(x) + f(y))\sqrt{P(x, y)} dy \right]^2 dx \\ &\leq \frac{1}{4} \int_\Omega \mu(x) \left[\int_\Omega P(x, y)(f(x) + f(y))^2 dy \int_\Omega P(x, y) dy \right] dx \\ &= \frac{1}{4} \int_\Omega \mu(x) \int_\Omega P(x, y)(f(x) + f(y))^2 dy dx\end{aligned}$$

Therefore

$$\text{Var}_\mu(f) - \text{Var}_\mu(P_{ZZ}f) \geq \frac{1}{4} \int_\Omega \int_\Omega \mu(x)P(x, y)(f(x) - f(y))^2 dx dy = \frac{1}{2}\mathcal{E}_\mu(f, f)$$

To get from variance to the L_2 -distance, notice that $\|\pi - \mu\|_2 = \text{Var}_\mu(\frac{\pi - \mu}{\mu}) = \text{Var}_\mu(\frac{\pi}{\mu})$ for any distribution π . Let $f_t(x) := \mu_t(x)/\mu(x)$. Then

$$\begin{aligned} P_{ZZ}f_t(x) &= \int_{\Omega} P_{ZZ}(x, y)f(y) dy = \int_{\Omega} P_{ZZ}(x, y)\frac{\mu_t(y)}{\mu(y)} dy \\ &= \int_{\Omega} \mu_t(y)\frac{P_{ZZ}(y, x)}{\mu(x)} dy = \frac{\mu_{t+1}(x)}{\mu(x)} = f_{t+1}(x) \end{aligned}$$

where the third equality comes from the reversibility of the Markov chain: $\mu(x)P_{ZZ}(x, y) = \mu(y)P_{ZZ}(y, x)$.

Finally, substituting f_t in (2):

$$\|\mu_{t+1} - \mu\|_2 = \text{Var}_\mu(f_{t+1}) = \text{Var}_\mu(P_{ZZ}f_t) \leq \left(1 - \frac{\lambda}{2}\right) \text{Var}_\mu(f_t) = \left(1 - \frac{\lambda}{2}\right) \|\mu_t - \mu\|_2$$

■

We conclude this section with the definition of conductance. Intuitively, it is the likelihood of escaping from the worst-case set in one step.

Definition 8 *The conductance of Markov chain $\mathcal{M} = (\Omega, P)$ with stationary distribution μ is*

$$\Phi_{\mathcal{M}} := \inf_{S \subseteq \Omega: 0 < \mu(S) < 1/2} \Pr(X_{t+1} \notin S \mid X_t \sim \mu(S))$$

The conductance can be viewed as a special case of the Poincaré inequality. If we restrict functions f to indicator functions, then $\lambda' := \inf_{f: \Omega \rightarrow \{0,1\}} \frac{\mathcal{E}_\mu(f, f)}{\text{Var}_\mu(f)}$ equals $\Phi_{\mathcal{M}}$. This points to a trivial relationship between the two quantities: $\lambda_{\mathcal{M}} \leq \Phi_{\mathcal{M}}$. Jerrum and Sinclair [9] proved a much stronger and more useful Cheeger-type inequality: $\lambda_{\mathcal{M}} \geq \Phi_{\mathcal{M}}^2/2$. Thus, obtaining a lower bound on the conductance provides a lower bound on the Poincaré constant.

3 Markov Chains for Sampling Points in Convex Bodies

The input for our algorithms is a convex body $K \subseteq \mathbb{R}^n$ of diameter D , and a precision ε (the bound on total variation distance from the uniform distribution). The body K is given by a *membership oracle*, i.e. if x is in K the answer is YES, otherwise it is NO. (In fact, because of the volume application, the body is usually given by a *separation oracle*, i.e. if $x \in K$, the answer is YES, otherwise the answer is a hyperplane separating x from K . This allows transformations of the body, e.g. the use of the ellipsoid method, or the isotropic position. A standard application of the ellipsoid method can transform K into a convex body containing the unit ball, of diameter $O(n^{3/2})$, and of the same volume as K .)

We will sample points using a ball-walk Markov chain approach by Lovász and Simonovits [15]. We denote δ the *step size* of the walk: Let $x \in K$ be the current state of the chain and $Y \subseteq K$ be the set of all points distant at most δ from x . Pick y , the new state of the chain, uniformly from Y . This describes our *speedy walk* on K . More precisely,

Definition 9 (Speedy Walk.) Fix $\delta > 0$ and denote $B(x, r)$ the n -dimensional ball of radius r centered at x . If x is the current state of the speedy walk, then the next state y is chosen uniformly at random from $B(x, \delta) \cap K$. In other words, the transition distribution $P(x, y) = \frac{1}{\text{vol}_n(B(x, \delta) \cap K)}$ for $\|x - y\| \leq \delta$, and $P(x, y) = 0$ otherwise.

As we will see shortly, the stationary distribution of the speedy walk is not the uniform distribution. Fortunately, a little trick can get us from speedy to the uniform distribution. This will be discussed in section 5.5.

Another problem with the speedy walk is its implementation in practice. To be able to simulate one step of this walk, we need to sample from $B(x, \delta) \cap K$. We will sample y' uniformly from $B(x, \delta)$ and if y' happens to be in K , we will move to $y := y'$, otherwise we stay in x , i.e. $y := x$. This type of walk we call the *Metropolis walk* (a shortcut for “Metropolis version of the ball walk”). Simulating one step of the speedy walk may require many steps of the Metropolis walk (the expected number of Metropolis steps is $\frac{\text{vol}_n B(0, \delta)}{\text{vol}_n(B(x, \delta) \cap K)}$). Again, we will overcome this problem in section 5.3.

We will be able to analyze the mixing time of the speedy walk using the Poincaré inequality (section 4). However, first we need to know the stationary distribution of this walk. Intuitively, it should be proportional to $B(x, \delta) \cap K$. Before we make this intuition formal, we define the local conductance, i.e. percentage of points in δ -ball centered at X which are in K :

Definition 10 For $x \in K$ let $l(x) := \frac{\text{vol}_n(B(x, \delta) \cap K)}{\text{vol}_n B(0, \delta)}$. The function $l : K \rightarrow \mathbb{R}$ is called the local conductance on K .

We claim that the stationary distribution $\tilde{\mu}$ of the speedy walk is

$$\tilde{\mu}(x) = l(x)/L, \quad \text{where} \quad L = \int_K l(y) dy.$$

We need to verify that $\tilde{\mu}P = \tilde{\mu}$, where P is the transition function of the speedy walk.

$$\begin{aligned} (\tilde{\mu}P)(x) &= \int_{y \in K} \tilde{\mu}(y)P(y, x) dy = \int_{y \in B(x, \delta) \cap K} \tilde{\mu}(y)P(y, x) dy = \\ &= \int_{y \in B(x, \delta) \cap K} \frac{l(y)}{L} \cdot \frac{1}{\text{vol}_n(B(y, \delta) \cap K)} dy = \\ &= \frac{1}{L} \int_{y \in B(x, \delta) \cap K} \frac{1}{\text{vol}_n B(0, \delta)} dy = \frac{l(x)}{L} = \tilde{\mu}(x) \end{aligned}$$

Throughout this text, $\tilde{\mu}$ denotes the **stationary distribution of the speedy walk** (and the Metropolis walk), and μ is the **uniform distribution** on K .

Note 11 We have not proven that $\tilde{\mu}$ is the unique stationary distribution. This will follow from the Poincaré inequality for the speedy walk (Section 4) and from Theorem 6.

4 Poincaré Inequality

This section is devoted to the proof of the Poincaré inequality of the speedy walk. We shall follow the proof of Jerrum for convex bodies satisfying the curvature condition (“rounded” bodies, see

[8]). Our proof extends his proof for general convex bodies and is based on many hints from Mark Jerrum himself.

Theorem 12 (Poincaré inequality) *Let $K \subseteq \mathbb{R}^n$ be a convex body of diameter D and let $\delta \leq c_\delta D/\sqrt{n}$ (where c_δ is independent of n). For any (measurable) function $f : K \rightarrow \mathbb{R}$,*

$$\mathcal{E}_{\tilde{\mu}}(f, f) \geq \lambda \text{Var}_{\tilde{\mu}} f,$$

where

$$\lambda = \frac{c_\lambda \delta^2}{D^2 n}$$

for some dimension-independent constant c_λ .

Throughout this section we compute expectation, variance, and Dirichlet form w.r.t. $\tilde{\mu}$. To simplify notation, if the subscript $\tilde{\mu}$ is clearly understood from the context, we omit it.

We will find very helpful special notation for the expected value, variance, and the Dirichlet form restricted to only part of the original domain (a subset of K), while the distribution $\tilde{\mu}$ remains to be defined on the whole convex body K .

Notation 13 *Let $f : K \rightarrow \mathbb{R}$ and $K' \subseteq K$. We write*

$$\begin{aligned} E_{K'} f &= \int_{K'} \tilde{\mu}(x) f(x) \, dx, \\ \text{Var}_{K'} f &= \int_{K'} \tilde{\mu}(x) \left(f(x) - \frac{1}{\tilde{\mu}(K')} E_{K'} f \right)^2 \, dx, \\ \mathcal{E}_{K'} f &= \int_{K'} \tilde{\mu}(x) h(x) \, dx. \end{aligned}$$

4.1 Needle-like Body

In this subsection we prove that if the Poincaré inequality is violated for the whole body, then there exists a needle-like body (a body whose all but one dimensions are bounded by a given ϵ) violating the Poincaré inequality.

Lemma 14 *If there exists $f : K \rightarrow \mathbb{R}$ violating the Poincaré inequality, i.e. $\mathcal{E}_{\tilde{\mu}}(f, f) < \lambda \text{Var}_{\tilde{\mu}} f$, then for every $\epsilon > 0$, there exists a convex subset $K_1 \subseteq K$ such that $K_1 \subseteq [0, D] \times [0, \epsilon]^{n-1}$ in some Cartesian coordinate system, and*

$$\mathcal{E}_{K_1}(f, f) < \lambda \text{Var}_{K_1} f,$$

while

$$E_{K_1} f = 0.$$

The idea of the proof is simple: We will eliminate the “fat” dimensions one by one. As it turns out, to do this we need to have at least two “fat” dimensions. Then we are able to eliminate one of them, using these two geometric statements:

Theorem 15 *Let S be a convex set in \mathbb{R}^2 . Then there exists a point x such that any line going through x cuts S into two parts, area of each of which is at least $4/9$ of the area of S .*

For the proof, see e.g. [6, page 118]. The constant $4/9$ is tight but for our purposes it could be replaced by any positive number – for an easy proof when the constant is $1/3$, see [8, Lemma 6.11].

Lemma 16 *Let S be a convex set in \mathbb{R}^2 . Then $S \subseteq [0, \sqrt{2 \text{ area}(S)}] \times \mathbb{R}$ in some Cartesian coordinate system.*

Proof :

Let d be the diameter of S and let $A, B \in S$ be two points distant d from each other, i.e. $\|A-B\| = d$. The x -axis in our new coordinate system will be identical to line \overrightarrow{AB} . Let the y -axis be any line perpendicular to \overrightarrow{AB} . Let $C \in S$ be a point with the largest (non-negative) y -coordinate, and $D \in S$ be a point with the smallest (non-positive) y -coordinate. Then by the triangle inequality, $y_C - y_D \leq \|C - D\|$, and $\|C - D\| \leq d$ because d is the diameter. Also from the convexity of S we can conclude that the triangles ABC and ABD are subsets of S . Then

$$\text{area}(S) \geq \text{area}(ABC) + \text{area}(ABD) = \frac{d}{2}(y_C - y_D) \geq \frac{(y_C - y_D)^2}{2}$$

Therefore $y_C - y_D \leq \sqrt{2 \text{ area}(S)}$. ■

We are now ready for the proof of Lemma 14.

Proof of Lemma 14 :

Elimination of the “fat” dimensions is based on the following claim:

CLAIM: *Let $i \geq 2$, $K_i \subseteq K$ such that in some Cartesian coordinate system $K_i \subseteq [0, D]^i \times [0, \epsilon]^{n-i}$, and f violates the Poincaré inequality on K_i , i.e. $\mathcal{E}_{K_i}(f, f) < \lambda \text{Var}_{K_i} f$ and $E_{K_i} f = 0$. Then there exists $K_{i-1} \subseteq K_i$ such that in some Cartesian coordinate system $K_{i-1} \subseteq [0, D]^{i-1} \times [0, \epsilon]^{n-i+1}$, and f violates the Poincaré inequality on K_{i-1} .*

PROOF: Let L be the projection of K_i onto the i -th and $(i-1)$ -st dimensions (the last two “fat” dimensions). By Theorem 15 there exists a point $x_L \in L$ such that any line going through x_L cuts L into two “large” pieces. Let H be any hyperplane in \mathbb{R}^n containing x_L and orthogonal to the i -th and $(i-1)$ -st dimensions. This hyperplane cuts K_i into two parts $K_i \cap H^+$ and $K_i \cap H^-$. Our goal is to show that there exists H_* such that $E_{K_i \cap H_*^+} = E_{K_i \cap H_*^-} = 0$. Clearly,

$$\begin{aligned} \mathcal{E}_{K_i}(f, f) &= \mathcal{E}_{K_i \cap H_*^+}(f, f) + \mathcal{E}_{K_i \cap H_*^-}(f, f), \\ \text{Var}_{K_i} f &= \text{Var}_{K_i \cap H_*^+} f + \text{Var}_{K_i \cap H_*^-} f, \end{aligned}$$

and by the assumption

$$\mathcal{E}_{K_i}(f, f) < \lambda \text{Var}_{K_i} f.$$

If $E_{K_i \cap H_*^+} = E_{K_i \cap H_*^-} = 0$, then either $K_i \cap H_*^+$, or $K_i \cap H_*^-$ violates the Poincaré inequality. W.l.o.g. let it be $K_i \cap H_*^+$. By Theorem 15 we know that the area of $L \cap H_*^+$ (the projection of $K_i \cap H_*^+$ onto the i -th and $(i-1)$ -st dimensions) is at most $5/9$ of the area of L . Now we can take $K_i = K_i \cap H_*^+$ and $L = L \cap H_*^+$. By continuing this process and cutting K_i sufficiently many times we can guarantee that the area of L becomes smaller than $\frac{1}{2}\epsilon^2$. Then, by Lemma 16 there exists Cartesian coordinate system in which $L \subseteq \mathbb{R} \times [0, \epsilon]$. Thus, the new K_i is in fact $K_{i-1} \subseteq [0, D]^{i-1} \times [0, \epsilon]^{n-i+1}$.

It remains to prove the existence of H_* . By the assumption, $E_{K_i} f = E_{K_i \cap H^+} f + E_{K_i \cap H^-} f = 0$. W.l.o.g. let $E_{K_i \cap H^+} f \geq 0$. By rotating H around x_L by π we get H_0 such that $E_{K_i \cap H_0^+} f \leq 0$. Thus, by continuity there exists H^* satisfying $E_{K_i \cap H_*^+} f = 0$. ■

4.2 Shrinking the Last Dimension

Lemma 17 *Let K_1 be the subset of K obtained in Lemma 14. Let D' be such that $K_1 \subseteq [0, D'] \times [0, \epsilon]^{n-1}$ and there exists a point in K_1 with the first coordinate equal to 0 and a point with the first coordinate equal to D' . (In other words, K_1 fits tightly in this box w.r.t. the first coordinate.) Then for every sufficiently small $\epsilon > 0$ and every constant $c_\eta > 0$ such that $c_\eta c_\delta \leq \frac{4}{100}$, there exists a subset $K_0 \subseteq K_1$ such that $K_0 \subseteq [0, \eta] \times [0, \epsilon]^{n-1}$, for $\eta := \frac{c_\eta \delta}{\sqrt{n}} \cdot \frac{D'}{D}$, in some Cartesian coordinate system, and*

$$\mathcal{E}_{K_0}(f, f) < c_0 \text{Var}_{K_0} f,$$

where c_0 is a constant inversely proportional to c_η^2 .

Notice that in the last lemma we do not have any conditions on the expected value.

Proof of Lemma 17 (part 1):

We will show that we can chop K_1 by cuts perpendicular to the long axis into m measurable sets S_0, \dots, S_{m-1} , each of which fits in a box of size $[0, \eta/2] \times [0, \epsilon]^{n-1}$, such that at least one of the S_i or the $S_i \cup S_{i+1}$ will be a suitable candidate for K_0 .

Ideally, if we were able to write $\text{Var}_{K_1} f \leq \frac{c_0}{\lambda} \sum_{i=0}^{m-1} \text{Var}_{S_i} f$, then we would get:

$$\sum_{i=0}^{m-1} \mathcal{E}_{S_i}(f, f) = \mathcal{E}_{K_1}(f, f) < \lambda \text{Var}_{K_1} f \leq c_0 \sum_{i=0}^{m-1} \text{Var}_{S_i} f, \quad (3)$$

where the first equality follows from additivity of integrals, and the next inequality is given by the assumption of the lemma. Then by the pigeon-hole principle there would exist i such that

$$\mathcal{E}_{S_i}(f, f) < c_0 \text{Var}_{S_i} f$$

What is the sum $\sum_{i=0}^{m-1} \text{Var}_{S_i} f$ equal to? For simplicity, for a measurable set $S \subseteq K$ let us denote $\bar{f}_S := \frac{1}{\tilde{\mu}(S)} \mathbb{E}_S f$, in other words, \bar{f}_S is the actual expected value of f with the distribution $\tilde{\mu}$ properly restricted on the set S .

$$\begin{aligned} \sum_{i=0}^{m-1} \text{Var}_{S_i} f &= \sum_{i=0}^{m-1} \int_{S_i} \tilde{\mu}(x) [(f(x) - \bar{f}_{K_1}) - (\bar{f}_{S_i} - \bar{f}_{K_1})]^2 dx \\ &= \sum_{i=0}^{m-1} \int_{S_i} \tilde{\mu}(x) (f(x) - \bar{f}_{K_1})^2 dx + \sum_{i=0}^{m-1} \int_{S_i} \tilde{\mu}(x) (\bar{f}_{S_i} - \bar{f}_{K_1})^2 dx - \\ &\quad - 2 \sum_{i=0}^{m-1} \int_{S_i} \tilde{\mu}(x) (f(x) - \bar{f}_{K_1}) (\bar{f}_{S_i} - \bar{f}_{K_1}) dx \\ &= \text{Var}_{K_1} f + \sum_{i=0}^{m-1} \tilde{\mu}(S_i) (\bar{f}_{S_i} - \bar{f}_{K_1})^2 - 2 \sum_{i=0}^{m-1} (\bar{f}_{S_i} - \bar{f}_{K_1}) (\mathbb{E}_{S_i} f - \tilde{\mu}(S_i) \bar{f}_{K_1}) \\ &= \text{Var}_{K_1} f - \sum_{i=0}^{m-1} \tilde{\mu}(S_i) (\bar{f}_{S_i} - \bar{f}_{K_1})^2, \end{aligned}$$

where the last equality follows from $\tilde{\mu}(S) \bar{f}_S = \mathbb{E}_S f$. In other words, this calculation tells us that the variance restricted to K_1 can be written as a sum of variances within each S_i plus the variance

between all of these sets:

$$\text{Var}_{K_1} f = \sum_{i=0}^{m-1} \text{Var}_{S_i} f + \sum_{i=0}^{m-1} \tilde{\mu}(S_i) (\bar{f}_{S_i} - \bar{f}_{K_1})^2 \quad (4)$$

If $\sum_{i=0}^{m-1} \text{Var}_{S_i} f \geq \sum_{i=0}^{m-1} \tilde{\mu}(S_i) (\bar{f}_{S_i} - \bar{f}_{K_1})^2$, then we have $\text{Var}_{K_1} f \leq 2 \sum_{i=0}^{m-1} \text{Var}_{S_i} f$. Since we may assume that $2 \leq c_0/\lambda$, e.g. by setting $c_0 \geq 2c_\eta$ (recall that $\delta < D$), this case is resolved, see 3.

If $\sum_{i=0}^{m-1} \text{Var}_{S_i} f < \sum_{i=0}^{m-1} \tilde{\mu}(S_i) (\bar{f}_{S_i} - \bar{f}_{K_1})^2$, the analysis is more complicated. We will prove that in this case one of the $U_i := S_i \cup S_{i+1}$ (for properly defined S_i) satisfies the desired inequality. As in the first case, we will want to express $\text{Var}_{K_1} f$ in terms of $\sum_{i=0}^{m-2} \text{Var}_{U_i} f$. Clearly, we assume

$$\text{Var}_{K_1} f < 2 \sum_{i=0}^{m-1} \tilde{\mu}(S_i) (\bar{f}_{S_i} - \bar{f}_{K_1})^2 \quad (5)$$

To simplify notation, we write $w_i = \tilde{\mu}(S_i)$. Also, notice that

$$\bar{f}_{U_i} = \frac{w_i \bar{f}_{S_i} + w_{i+1} \bar{f}_{S_{i+1}}}{w_i + w_{i+1}}$$

Applying (4) on U_i instead of K_1 , we get:

$$\begin{aligned} \text{Var}_{U_i} f &= \sum_{k=i}^{i+1} \text{Var}_{S_k} f + \sum_{k=i}^{i+1} w_k (\bar{f}_{S_k} - \bar{f}_{U_i})^2 \\ &\geq w_i (\bar{f}_{S_i} - \bar{f}_{U_i})^2 + w_{i+1} (\bar{f}_{S_{i+1}} - \bar{f}_{U_i})^2 \\ &= w_i \left[\frac{w_{i+1}}{w_i + w_{i+1}} (\bar{f}_{S_i} - \bar{f}_{S_{i+1}}) \right]^2 + w_{i+1} \left[\frac{w_i}{w_i + w_{i+1}} (\bar{f}_{S_{i+1}} - \bar{f}_{S_i}) \right]^2 \\ &= \frac{w_i w_{i+1}}{w_i + w_{i+1}} (\bar{f}_{S_i} - \bar{f}_{S_{i+1}})^2 \end{aligned}$$

We will be able to express $\text{Var}_{K_1} f$ in terms of $\frac{w_i w_{i+1}}{w_i + w_{i+1}} (\bar{f}_{S_i} - \bar{f}_{S_{i+1}})^2$.

If we divide inequality (5) by $\tilde{\mu}(K_1)$ the right hand side is a variance of a discrete random variable $X : \{0, \dots, m-1\} \rightarrow \mathbb{R}$, where $X_i := \bar{f}_{S_i}$ with distribution $\pi_i = \tilde{\mu}(S_i)/\tilde{\mu}(K_1)$. (Notice that $E_\pi(X) = 0$, because $E_{K_1} f = 0$.)

$$\begin{aligned} \text{Var}_\pi X &= \sum_{i=0}^{m-1} \pi_i X_i^2 - (E_\pi X)^2 \\ &= \frac{1}{2} \sum_{i=0}^{m-1} \pi_i X_i^2 + \frac{1}{2} \sum_{j=0}^{m-1} \pi_j X_j^2 - \sum_{i=0}^{m-1} \pi_i X_i \sum_{j=0}^{m-1} \pi_j X_j \\ &= \frac{1}{2} \sum_{0 \leq i < j < m} \pi_i \pi_j (X_i^2 + X_j^2 - 2X_i X_j) \\ &= \frac{1}{2} \sum_{0 \leq i < j < m} \pi_i \pi_j (X_i - X_j)^2 \end{aligned}$$

Therefore,

$$\frac{1}{\tilde{\mu}(K_1)} \text{Var}_{K_1} f < 2 \text{Var}_{\pi} X = \sum_{0 \leq i < j < m} \pi_i \pi_j (X_i - X_j)^2$$

Or, equivalently, we get

$$\begin{aligned} \tilde{\mu}(K_1) \text{Var}_{K_1} f &< \sum_{0 \leq i < j < m} w_i w_j (\bar{f}_{S_i} - \bar{f}_{S_j})^2 \\ &= \sum_{0 \leq i < j < m} w_i w_j \left[\sum_{k=i}^{j-1} \sqrt{\frac{w_k + w_{k+1}}{w_k w_{k+1}}} \cdot \sqrt{\frac{w_k w_{k+1}}{w_k + w_{k+1}}} (\bar{f}_{S_k} - \bar{f}_{S_{k+1}}) \right]^2 \\ &\leq \sum_{0 \leq i < j < m} w_i w_j \left[\sum_{k=i}^{j-1} \frac{w_k + w_{k+1}}{w_k w_{k+1}} \right] \cdot \left[\sum_{k=i}^{j-1} \frac{w_k w_{k+1}}{w_k + w_{k+1}} (\bar{f}_{S_k} - \bar{f}_{S_{k+1}})^2 \right], \end{aligned}$$

where in the bound we applied the Cauchy-Schwarz inequality. Let us denote

$$a_{i,j} := w_i w_j \sum_{k=i}^{j-1} \frac{w_k + w_{k+1}}{w_k w_{k+1}} \quad (6)$$

Then

$$\begin{aligned} \tilde{\mu}(K_1) \text{Var}_{K_1} f &< \sum_{0 \leq i < j < m} a_{i,j} \sum_{k=i}^{j-1} \frac{w_k w_{k+1}}{w_k + w_{k+1}} (\bar{f}_{S_k} - \bar{f}_{S_{k+1}})^2 \\ &\leq \sum_{0 \leq i < j < m} a_{i,j} \sum_{k=i}^{j-1} \text{Var}_{U_k} f \end{aligned}$$

Similarly to the first case, if we were able to show that $\sum_{i < j} a_{i,j} \leq \frac{c_0}{\lambda} \cdot \frac{\tilde{\mu}(K_1)}{2}$, see (3), then we would get:

$$\sum_{i=0}^{m-2} \mathcal{E}_{U_i}(f, f) \leq 2 \mathcal{E}_{K_1}(f, f) < 2 \lambda \text{Var}_{K_1} f \leq c_0 \sum_{i=0}^{m-1} \text{Var}_{U_i} f, \quad (7)$$

allowing us to conclude that there exists i such that

$$\mathcal{E}_{U_i}(f, f) < c_0 \text{Var}_{U_i} f$$

■

In the rest of this subsection we will bound $\sum_{i < j} a_{i,j}$. Notice that until now the only important property of the S_i was that they fit in a sufficiently small box. For bounding the $\sum_{i < j} a_{i,j}$ we will need something more:

Definition 18 *Let us fix the Cartesian coordinate system for $K_1 \subseteq [0, D'] \times [0, \epsilon]^{n-1}$ from Lemma 14, $D' \leq D$. For $x \in \mathbb{R}$ we denote $S(x) := \{A \in K_1 \mid A^{(1)} = x\}$, where $A^{(i)}$ means the i -th coordinate of A , and $l_{\max}(x) := \max_{A \in S(x)} l(A)$. In other words, l_{\max} is the maximal l over all points in K_1 with fixed first coordinate. Notice that $l_{\max}(0) > 0$ and $l_{\max}(D') > 0$. We inductively define x_i :*

- $x_0 := 0$,
- For $i > 0$, let x_i be the largest number satisfying: $\frac{1}{2} \leq \frac{l(x_i)}{l(x_{i-1})} \leq 2$, and $|x_i - x_{i-1}| \leq \frac{\eta}{2}$.

We define slabs $S_i := K_1 \cap ([x_i, x_{i+1}] \times [0, \epsilon]^{n-1})$ and denote $\text{width}(S_i) := |x_{i+1} - x_i|$. Let m be the total number of slabs.

First we study basic properties of the S_i .

Lemma 19 *For the above definition of the S_i there exist l, r such that $|x_{i+1} - x_i| = \frac{\eta}{2}$ if and only if $i \in \{l, \dots, r-1\}$. Furthermore, for $i < l$ the sequence $\{|x_{i+1} - x_i|\}_{i=0}^{l-1}$ is non-decreasing and for $i \geq r$ the sequence $\{|x_{i+1} - x_i|\}_{i=r}^{m-1}$ is non-increasing.*

Before we start with the proof of this lemma, we need to explore some properties of the local conductance l . For two sets $L_1, L_2 \subseteq \mathbb{R}^n$ we define their *Minkowski sum* $L_1 + L_2 := \{z_1 + z_2 \mid z_1 \in L_1, z_2 \in L_2\}$. Now we can state the Brunn-Minkowski Theorem (see e.g. [6, Chapter 26, Thm 46]), a helpful tool when proving geometric statements.

Theorem 20 (Brunn-Minkowski) *Let K' and K'' be two convex bodies in \mathbb{R}^n . Then*

$$\text{vol}_n(K' + K'')^{1/n} \geq \text{vol}_n(K')^{1/n} + \text{vol}_n(K'')^{1/n}$$

The following lemma is the heart of several proofs in this work.

Lemma 21 *Let K and L be two convex bodies in \mathbb{R}^n and let $g(u) = \text{vol}_n(K \cap (u + L))$. We define the domain of g to be $D_g := \{u \mid g(u) > 0\}$. Then the function $g(u)^{1/n}$ is concave and $g(u)$ is log-concave over the domain D_g .*

Proof :

First we prove that $g(u)^{1/n}$ is concave. Let $\kappa \in [0, 1]$, $x, y \in \mathbb{R}^n$. We want to show that $g(\kappa x + (1 - \kappa)y) \geq \kappa g(x) + (1 - \kappa)g(y)$. Suppose we know that $K \cap (\kappa x + (1 - \kappa)y + L) \supseteq \kappa(K \cap (x + L)) + (1 - \kappa)(K \cap (y + L))$. Then:

$$\begin{aligned} g(\kappa x + (1 - \kappa)y)^{1/n} &= \text{vol}_n[K \cap (\kappa x + (1 - \kappa)y + L)]^{1/n} \\ &\geq \text{vol}_n[\kappa(K \cap (x + L)) + (1 - \kappa)(K \cap (y + L))]^{1/n} \\ &\geq \text{vol}_n[\kappa(K \cap (x + L))]^{1/n} + \text{vol}_n[(1 - \kappa)(K \cap (y + L))]^{1/n} \\ &= (\kappa^n)^{1/n} \text{vol}_n[(K \cap (x + L))]^{1/n} + ((1 - \kappa)^n)^{1/n} \text{vol}_n[(K \cap (y + L))]^{1/n} \\ &= \kappa \text{vol}_n[(K \cap (x + L))]^{1/n} + (1 - \kappa) \text{vol}_n[(K \cap (y + L))]^{1/n} \\ &= \kappa g(x)^{1/n} + (1 - \kappa) g(y)^{1/n}, \end{aligned}$$

where the second inequality follows from the Brunn-Minkowski Theorem. This implies that $g(u)^{1/n}$ is concave and since the composition of two concave functions is a concave function, $\log(g(u)^{1/n}) = \frac{1}{n} \log g(u)$ is also concave. Thus, $g(u)$ is log-concave.

It remains to prove that $K \cap (\kappa x + (1 - \kappa)y + L) \supseteq \kappa(K \cap (x + L)) + (1 - \kappa)(K \cap (y + L))$. Let $a \in \kappa(K \cap (x + L))$, $b \in (1 - \kappa)(K \cap (y + L))$. We want to show that $a + b \in K \cap (\kappa x + (1 - \kappa)y + L)$. By the assumption, $a \in \kappa K$, $b \in (1 - \kappa)K$, therefore $a + b \in \kappa K + (1 - \kappa)K$. Since K is convex,

$\kappa K + (1 - \kappa)K = K$, thus $a + b \in K$. Also, $a \in \kappa(x + L) = \kappa x + \kappa L$, $b \in (1 - \kappa)(y + L) = (1 - \kappa)y + (1 - \kappa)L$, implying $a + b \in \kappa x + (1 - \kappa)y + L$. This finishes the proof of the lemma. ■

These are the main properties of the local conductance l :

Lemma 22 *The local conductance l satisfies:*

1. $l(x)^{1/n}$ is concave over K ,
2. $l(x)$ is log-concave over K ,
3. $l_{\max}(x)^{1/n}$ is concave and $l_{\max}(x)$ is log-concave over K ,
4. $\ln l(x)$ is Lipschitz over K with constant n/δ , i.e. for $x, y \in K$, $|\ln l(x) - \ln l(y)| \leq \frac{n}{\delta} \|x - y\|$

Proof :

Items 1 and 2 follow directly from previous lemma.

For part 3, we will show that if $g : K_1 \rightarrow \mathbb{R}$ is a concave function, then $g_{\max}(x) := \max_{y \in S(x)} g(y)$ is concave over K_1 , too. This will imply that both $l_{\max}(x)^{1/n} = \max_{y \in S(x)} l(y)^{1/n}$, and $\log l_{\max}(x) = \max_{y \in S(x)} \log l(y)$ are concave by the first two parts of this theorem.

We want to show that for $\kappa \in [0, 1]$ and $x, y \in [0, D']$

$$g_{\max}(\kappa x + (1 - \kappa)y) \geq \kappa g_{\max}(x) + (1 - \kappa)g_{\max}(y)$$

Clearly, $g_{\max}(x)$ is achieved for some $A \in S(x)$, similarly $g_{\max}(y) = l(B)$ for some $B \in S(y)$. Then by concavity of $g(x)$ we have $g(\kappa A + (1 - \kappa)B) \geq \kappa g(A) + (1 - \kappa)g(B)$. One can easily verify that $\kappa A + (1 - \kappa)B \in S(\kappa x + (1 - \kappa)y)$ (the body K_1 is convex, therefore $\kappa A + (1 - \kappa)B \in K_1$).

$$\begin{aligned} g_{\max}(\kappa x + (1 - \kappa)y) &\geq g(\kappa A + (1 - \kappa)B) \\ &\geq \kappa g(A) + (1 - \kappa)g(B) \\ &= \kappa g_{\max}(x) + (1 - \kappa)g_{\max}(y) \end{aligned}$$

For case 4, notice that the definition of $l(x)$ makes sense even for x outside of K . By Lemma 21, $l(x)^{1/n}$ is concave over D_l . It is easy to see that the domain $D_l = K + B(0, \delta)$. Let $A, B \in K$ and let C be on the halfline \overrightarrow{AB} , distant δ from B and $\|A - B\| + \delta$ from A . Clearly, $C \in D_l$. Then by concavity,

$$l(B)^{1/n} \geq \frac{\delta}{\|A - B\| + \delta} l(A)^{1/n} + \frac{\|A - B\|}{\|A - B\| + \delta} l(C)^{1/n} \geq \frac{\delta}{\|A - B\| + \delta} l(A)^{1/n}$$

Then

$$\left[\frac{l(A)}{l(B)} \right]^{1/n} \leq \frac{\|A - B\| + \delta}{\delta}$$

Taking logarithm of both sides

$$|\ln l(A) - \ln l(B)| \leq n \ln \left(1 + \frac{\|A - B\|}{\delta} \right) \leq \frac{n}{\delta} \|A - B\|$$

■

Proof of Lemma 19:

Since l_{\max} is a log-concave function, there exists x'_1, x'_2 such that l_{\max} is strictly increasing for $x \in [0, x'_1]$, strictly decreasing for $x \in [x'_2, D']$, and constant for $x \in [x'_1, x'_2]$. Let $\text{width}(S_i) := |x_{i+1} - x_i|$. We will prove that $\text{width}(S_{i-1}) \leq \text{width}(S_i)$ for i such that $x_{i+1} \leq x'_1$. (In other words, while l_{\max} increases the width is non-decreasing).

Let i be such that $x_{i+1} \leq x'_1$ and suppose $\text{width}(S_i) < \text{width}(S_{i-1})$. This means that width of S_i is smaller than $\eta/2$ and by definition of the S_i , $l_{\max}(x_{i+1}) = 2l_{\max}(x_i)$, or equivalently $\log l_{\max}(x_{i+1}) = \log l_{\max}(x_i) + 1$. But $\log l_{\max}$ is a concave function:

$$\log l_{\max}(x_i) \geq \frac{1}{2}[\log l_{\max}(x_i - \text{width}(S_i)) + \log l_{\max}(x_{i+1})]$$

Putting the last two observations together we get $l_{\max}(x_i) \geq 2l_{\max}(x_i - \text{width}(S_i))$. Since $\text{width}(S_i) < \text{width}(S_{i-1})$, this means $l_{\max}(x_i) > 2l_{\max}(x_{i-1})$, a contradiction with the definition of the S_i .

Similar argument shows that when l_{\max} starts decreasing, the width is non-increasing. Therefore the slabs of “full” width $\eta/2$ are the last slabs while l_{\max} increases and the first slabs when l_{\max} decreases, proving the existence of l and r . ■

Now we know that the S_i can be divided into three groups: The “left” group, ($i < l$), then the “middle” group ($l \leq i < r$), and the “right” group ($r \leq i$). We will show that each of these groups has specific properties (the “left” and “right” group have similar) which will help us to bound the $\sum_{i < j} a_{i,j}$.

We split the sum into several smaller sums and we will be able to evaluate each of them separately.

$$\sum_{i < j} a_{i,j} = \sum_{i < j < l} a_{i,j} + \sum_{i < l \leq j < r} a_{i,j} + \sum_{i < l, r \leq j} a_{i,j} + \sum_{l \leq i < j < r} a_{i,j} + \sum_{l \leq i < r \leq j} a_{i,j} + \sum_{r \leq i < j} a_{i,j} \quad (8)$$

$\sum_{i < j < l}$ will be computed in essentially the same way as $\sum_{r \leq i < j}$, similarly $\sum_{i < l \leq j < r}$ and $\sum_{l \leq i \leq r \leq j}$.

We will first examine the slabs in the “middle” group with the help of a generalization of the Brunn-Minkowski Theorem, due to Dinghas (see [3, Thm 10.1]):

Theorem 23 (Dinghas) *Let A_1 and A_2 be two bounded measurable sets in \mathbb{R}^n , f be a non-negative measurable function defined on $A_1 \cup A_2$, and $p > 0$. Let*

$$g(x) := \sup\{(f(x_1)^{1/p} + f(x_2)^{1/p})^p \mid x = x_1 + x_2, x_i \in A_i\}$$

Then

$$\left[\int_{A_1 + A_2} g(x) dx \right]^{1/(p+n)} \geq \left[\int_{A_1} f(x) dx \right]^{1/(p+n)} + \left[\int_{A_2} f(x) dx \right]^{1/(p+n)}$$

Lemma 24 *The sequence $w_l^{1/(2n)}, \dots, w_{r-1}^{1/(2n)}$ is concave. Consequently, sequence $1/w_l, \dots, 1/w_{r-1}$ is convex.*

Proof :

Let $i \in \{l + 1, \dots, r - 2\}$. We need to prove that $2w_i^{1/(2n)} \geq w_{i-1}^{1/(2n)} + w_{i+1}^{1/(2n)}$. Since all $S_{i-1}, S_i,$

and S_{i+1} have the same width and K_1 is convex, it is easy to observe that $2S_i \supseteq S_{i-1} + S_{i+1}$. By Lemma 22, $l(x)^{1/n}$ is concave and thus if $x = x_1 + x_2$, we get

$$l\left(\frac{1}{2}(x_1 + x_2)\right)^{1/n} \geq \frac{1}{2}(l(x_1)^{1/n} + l(x_2)^{1/n}),$$

or equivalently

$$l\left(\frac{x}{2}\right) \geq \frac{1}{2^n} \sup\{(l(x_1)^{1/n} + l(x_2)^{1/n})^n \mid x = x_1 + x_2\}$$

Now we can apply Dinghas's Theorem for $A_1 := S_{i-1}$, $A_2 := S_{i+1}$, $f(x) := l(x)$, and $p := n$:

$$\begin{aligned} 2w_i^{1/(2n)} &= 2 \left[\int_{S_i} l(x) dx \right]^{1/(2n)} = 2 \left[\int_{2S_i} \frac{l(x/2)}{2^n} dx \right]^{1/(2n)} \\ &\geq 2 \left[\int_{2S_i} \frac{g(x)}{2^{2n}} dx \right]^{1/(2n)} \\ &\geq \left[\int_{S_{i-1}} l(x) dx \right]^{1/(2n)} + \left[\int_{S_{i+1}} l(x) dx \right]^{1/(2n)} \\ &= w_{i-1}^{1/(2n)} + w_{i+1}^{1/(2n)} \end{aligned}$$

This proves the concavity of $\{a_i\}_{i=l}^{r-1}$, where $a_i := w_i^{1/(2n)}$. If we apply function $h(x) := x^{-2n}$ to this sequence, we get

$$h(a_{i-1}) + h(a_{i+1}) \geq 2h\left(\frac{a_{i-1} + a_{i+1}}{2}\right) \geq 2h(a_i),$$

where the first inequality is convexity of $h(x)$, and the second inequality comes from the fact that $h(x)$ is a decreasing function and a_i is a concave sequence (and thus $(a_{i-1} + a_{i+1})/2 \leq a_i$). Therefore, the sequence $\{h(a_i)\}_{i=l}^{r-1}$ is convex. \blacksquare

The last lemma allows us to bound $\sum_{l \leq i < j < r} a_{i,j}$: Since the sequence $1/w_l, \dots, 1/w_r$ is convex, we have $\sum_{k=i}^j 1/w_k \leq (j - i + 1)(1/w_i + 1/w_j)/2$. Then for $l \leq i < j < r$,

$$a_{i,j} \leq 2w_i w_j \sum_{k=i}^j \frac{1}{w_k} \leq w_i w_j (j - i + 1) \left(\frac{1}{w_i} + \frac{1}{w_j} \right) = (j - i + 1)(w_j + w_i), \quad (9)$$

where the first inequality comes from the definition of $a_{i,j}$, see (6). Then

$$\begin{aligned} \sum_{l \leq i < j < r} a_{i,j} &\leq \sum_{l \leq i < j < r} (j - i + 1)(w_i + w_j) \\ &\leq (r - l) \sum_{l \leq i < j < r} (w_i + w_j) = (r - l) \left[\sum_{l \leq i < j < r} w_i + \sum_{l \leq i < j < r} w_j \right] \\ &= (r - l) \left[\sum_{l \leq i < r-1} (r - i - 1)w_i + \sum_{l < j < r} (j - l)w_j \right] \\ &\leq (r - l)(r - l - 1) \sum_{l \leq i < r} w_i \\ &\leq (r - l)^2 \tilde{\mu}(K_1) \end{aligned}$$

For convenience let us denote $M := r - l$, i.e. the number of slabs in the “middle” part. Thus

$$\sum_{l \leq i < j < r} a_{i,j} = O(M^2 \tilde{\mu}(K_1)) \quad (10)$$

Now we turn our attention to the “left” and “right” sections of the slabs. The number M can be easily bounded because we know the width of the slabs – thus we can bound $\sum_{l \leq i < j < r} a_{i,j}$ in terms of D , D' , δ , and n . We will use similar technique to bound $\sum_{i < j < l} a_{i,j}$ to get a bound in terms of l . Fortunately, we are able to lower-bound the width of slabs outside the “middle” section too.

Lemma 25 *For sufficiently small ϵ the width of any slab is at least $\frac{\delta}{2n}$.*

Proof :

This statement is clearly true for slabs in the “middle” section. For the “left” section (the proof for “right” section is similar), let $i < l$ and let $A \in S_i$ be a point with first coordinate x_i attaining $l_{\max}(x_i)$, and $B \in S_i$ with first coordinate x_{i+1} be such that $l(B) = l_{\max}(x_{i+1})$ (i.e. the leftmost and rightmost “maximal” points). Then, by Lipschitz, see Lemma 22, part 4, we get

$$\|A - B\| \geq \frac{\delta}{n} |\ln l(A) - \ln l(B)| = \frac{\delta}{n} \ln 2$$

Let A' be the point that we obtain from A by changing its first coordinate to x_{i+1} . Then the width of S_i is $\|A - A'\|$. By the triangle inequality we get a bound on the width of S_i : $\|A - A'\| \geq \frac{\delta}{n} \ln 2 - \epsilon \sqrt{n-1}$, where the term $\epsilon \sqrt{n-1}$ represents the upper-bound on the distance from B to A' . Remember that we are free to pick ϵ as small as we want, so we may assume $\text{width}(S_i) = \|A - A'\| \geq \frac{\delta}{2n}$. ■

If we were to use the same technique for evaluating $\sum_{i < j < l} a_{i,j}$ as we did for $\sum_{l \leq i < j < r} a_{i,j}$, we would get a bound in terms of l^2 . However, l^2 is more than we can afford. Therefore we need to know more about the “left” and “right” sections.

Lemma 26 *The sequence w_0, \dots, w_{l-1} is geometrically increasing and the sequence w_r, \dots, w_{m-1} is geometrically decreasing. Further, $w_{l-1} \leq c_w w_l$ and $w_r \leq c_w w_{r-1}$ for some constant $c_w > 0$.*

Proof :

First we prove that w_0, \dots, w_{l-1} is geometrically increasing (the proof for w_r, \dots, w_{m-1} is analogous). We want to find a constant $c_g > 1$ such that for any $i < l - 1$ we can show $c_g w_i \leq w_{i+1}$. Let $d := \text{width}(S_{i+1})$. We will prove that there exists a constant $c_g > 1$ such that $c_g \int_{S(x)} l(z) dz \leq \int_{S(x+d)} l(z) dz$ for any $x \in [x_i, x_{i+1}]$. Then

$$c_g w_i = \int_{x_i}^{x_{i+1}} c_g \int_{S(x)} l(z) dz dx \leq \int_{x_i}^{x_{i+1}} \int_{S(x+d)} l(z) dz dx \leq \int_{x_i+d}^{x_{i+1}+d} \int_{S(x)} l(z) dz dx \leq w_{i+1},$$

since $x_{i+1} + d = x_{i+2}$ and by Lemma 19 the $(i + 1)$ -st slab is at least as wide as the i -th slab (and thus $x_i + d \geq x_{i+1}$).

CLAIM: *Let $V(x) := \text{vol}_{n-1}(S(x))$. There exists a constant $c'_g > 1.1\sqrt{e}$ such that $c'_g V(x) l_{\max}(x) \leq V(x+d) l_{\max}(x+d)$ for every $x \in [x_i, x_{i+1}]$.*

Once we have proven this claim, it will suffice to relate $V(x) l_{\max}(x)$ to $\int_{S(x)} l(z) dz$.

PROOF: Since i is from the “left” section, it follows from log-concavity of l_{\max} that $2l_{\max}(x) \leq l_{\max}(x+d)$ (the same reasoning is used in the proof of Lemma 19, second paragraph). Therefore if $V(x) \leq V(x+d)$ the claim follows trivially.

The problem is if $V(x) > V(x+d)$. We show that $V(x)^{1/n}$ is a concave function and it will allow us to bound the speed of its decreasing.

We want to show that for $x, y \in [0, D']$ and $\kappa \in [0, 1]$

$$V(\kappa x + (1 - \kappa)y)^{1/n} \geq \kappa V(x)^{1/n} + (1 - \kappa)V(y)^{1/n} \quad (11)$$

For $\rho > 0$, let $S_\rho := [0, \rho] \times [0, \epsilon]^{n-1}$, and $g_\rho(z) := \text{vol}_n(K_1 \cap (z + S_\rho))^{1/n}$. Then $V(x) = \lim_{\rho \rightarrow 0} g_\rho(x, 0, \dots, 0)/\rho$. Out of convenience for $x \in \mathbb{R}$ we will shortcut $g_\rho(x, 0, \dots, 0)$ as $g_\rho(x)$. By Lemma 21 the function $g_\rho(z)^{1/n}$ is concave over K_1 . Therefore $g_\rho(\kappa x + (1 - \kappa)y)^{1/n} \geq \kappa g_\rho(x)^{1/n} + (1 - \kappa)g_\rho(y)^{1/n}$. Dividing both sides by $\rho^{1/n}$ and taking limit as ρ tends to 0 gives us (11), proving the concavity of $V(x)^{1/n}$.

We assume that $V(x)^{1/n} > V(x+d)^{1/n}$. Further, $V(D') \geq 0$ and $V(x)^{1/n}$ is concave. Therefore

$$V(x+d)^{1/n} \geq \frac{D' - (x+d)}{D' - x} V(x)^{1/n} + \frac{d}{D' - x} V(D')^{1/n} \geq \frac{D' - (x+d)}{D' - x} V(x)^{1/n}$$

Thus, $V(x+d)/V(x) \geq (1 - \frac{d}{D'-x})^n$. We will prove that since l_{\max} is increasing rapidly, $x \leq D'/2$. Also, $d \leq \frac{\eta}{2} = \frac{c_\eta \delta D'}{2\sqrt{n}D}$. Putting these together we get

$$\frac{V(x+d)}{V(x)} \geq \left(1 - \frac{\eta/2}{D'/2}\right)^n = \left(1 - \frac{c_\eta \delta}{\sqrt{n}D}\right)^n \geq \left(1 - \frac{c_\eta c_\delta}{n}\right)^n, \quad (12)$$

where the last inequality comes from $\delta \leq c_\delta D/\sqrt{n}$. Since $(1 - \frac{t}{n})^n \geq e^{-\frac{nt}{n-t}}$ for every $t \in \mathbb{R}$, and $c_\eta c_\delta < 1/2$ implies $n - c_\eta c_\delta > n/2$, we get $V(x+d) \geq e^{-2c_\eta c_\delta} V(x)$. Thus

$$\frac{2}{e^{2c_\eta c_\delta}} V(x) l_{\max}(x) \leq V(x+d) l_{\max}(x+d)$$

We need $c'_g := \frac{2}{e^{2c_\eta c_\delta}} > 1.1\sqrt{e}$ and this is satisfied if $c_\eta c_\delta \leq \frac{4}{100} < \frac{1}{2} \ln \frac{2}{1.1\sqrt{e}}$ (constant $\frac{4}{100}$ comes from the assumption of Lemma 17).

It remains to prove that $x \leq D'/2$. We know that $l_{\max}^{1/n}$ is concave, $l_{\max}(0) > 0$, and $2l_{\max}(x) \leq l_{\max}(x+d)$. Therefore

$$\frac{x}{x+d} l_{\max}(x+d)^{1/n} < \frac{d}{x+d} l_{\max}(0)^{1/n} + \frac{x}{x+d} l_{\max}(x+d)^{1/n} \leq l_{\max}(x)^{1/n} \leq \frac{1}{2^{1/n}} l_{\max}(x+d)^{1/n}$$

Dividing both sides by $l_{\max}(x+d)^{1/n}$ gives us $x < \frac{d}{2^{1/n}-1} \leq \frac{\eta}{2(2^{1/n}-1)}$. Now we substitute the expression for η and bound for δ : $x < \frac{c_\eta c_\delta D'}{n} \cdot \frac{1}{2(2^{1/n}-1)}$. The last step is to bound $1/(2^{1/n}-1)$. Since $e^t \geq 1+t$ for any real t , if $t = \frac{1}{n} \ln 2$ then $2^{1/n} \geq 1 + \frac{1}{n} \ln 2$ and thus $1/(2^{1/n}-1) \leq n/\ln 2$. Therefore our bound on x is $x < c_\eta c_\delta D'/(2 \ln 2)$. Now all we need is assumption that $c_\eta c_\delta \leq \ln 2$ which is clearly true since $c_\eta c_\delta \leq \frac{4}{100}$. We proved $x < D'/2$, closing the proof of the claim. \square

What is the relation of $V(x)l_{\max}(x)$ and $\int_{S(x)} l(z) dz$? Recall that by the Lipschitz inequality, Lemma 22, part 4, for any $y, z \in S(x)$ we have $|\ln l(y) - \ln l(z)| \leq \frac{n}{\delta} \epsilon \sqrt{n-1}$. In particular, if y is

such that $l(y) = l_{\max}(x)$ and ϵ is sufficiently small so that $\epsilon \frac{n\sqrt{n-1}}{\delta} < \frac{1}{2}$, then $l(z) \leq l_{\max}(x) \leq \sqrt{e} l(z)$ for any $z \in S(x)$. Thus

$$c'_g \int_{S(x)} l(z) dz \leq c'_g V(x) l_{\max}(x) \leq V(x+d) l_{\max}(x+d) \leq \int_{S(x+d)} \sqrt{e} l(z) dz$$

This gives us $c_g := c'_g / \sqrt{e} \geq 1.1$.

Finally, we want to prove the existence of the constant c_w . Let $x \in [x_{l-1}, x_l]$, and $d := \text{width}(S_{l-1})$. We know that $l_{\max}(x) \leq 4l_{\max}(x+d)$ (because l_{\max} varies by a factor of 2 within each S_i). If $V(x)$ is still increasing, we are fine with $c_w := 4\sqrt{e}$. If $V(x)$ already decreases, we may use analysis similar to previous case. The only thing that we cannot guarantee anymore is $x < D'/2$. But we know that $x_{l-1} < D'/2$, thus $x < (D' + \eta)/2$. Since $\eta \leq c_\eta c_\delta D'$ and $c_\eta c_\delta < 1/2$, for x this means: $x < \frac{3}{4}D'$. A computation similar to (12) leads to $V(x+d)/V(x) \geq e^{-4c_\eta c_\delta}$. Therefore in this case

$$\frac{1}{4e^{4c_\eta c_\delta}} V(x) l_{\max}(x) \leq V(x+d) l_{\max}(x+d)$$

To go from $V(x)l_{\max}(x)$ to w_{l-1} we can copy the same reasoning as before and we get $w_{l-1} \leq c_w w_l$ for $c_w := \frac{\sqrt{e}}{4e^{4c_\eta c_\delta}}$. This concludes the proof of this lemma. \blacksquare

With the help of the last lemma, we can estimate $\sum_{i < j < l} a_{i,j}$, $\sum_{i < l \leq j < r} a_{i,j}$, and $\sum_{i < l, r \leq j} a_{i,j}$. This will allow us to finish the proof of Lemma 17.

Caution: What follows is a tedious computation using lemmas proven above. We encourage readers to do the computations by themselves instead of trying to read this proof.

Proof of Lemma 17 (part 2):

Let $c_g > 1$ be the constant determining the speed of geometric increasing of w_0, \dots, w_{l-1} . (By the proof of Lemma 26 we can take $c_g := 1.1$.) First we bound $a_{i,j}$, assuming $i < j < l$:

$$a_{i,j} \leq 2w_i w_j \sum_{k=i}^j \frac{1}{w_k} \leq 2w_i w_j \sum_{k=i}^j \frac{1}{c_g^{k-i} w_i} = 2w_j \frac{1/c_g^{j-i} - 1}{1/c_g - 1} < 2\hat{c}_g w_j, \quad (13)$$

where $\hat{c}_g := \frac{1}{1-1/c_g}$. Then

$$\sum_{i < j < l} a_{i,j} \leq \sum_{i < j < l} 2\hat{c}_g w_j < 2\hat{c}_g \sum_{i < l} \hat{c}_g w_{l-1} = O(lw_{l-1}) \quad (14)$$

For $i < l \leq j < r$, we can estimate $a_{i,j}$ (using the same techniques as in (9) and (13)):

$$\begin{aligned} a_{i,j} &\leq 2w_i w_j \left[\sum_{k=i}^{l-1} \frac{1}{w_k} + \sum_{k=l}^j \frac{1}{w_k} \right] \\ &< 2w_i w_j \left[\hat{c}_g \frac{1}{w_i} + (j-l+1) \left(\frac{1}{w_l} + \frac{1}{w_j} \right) \right] \\ &\leq 2\hat{c}_g w_j + 2 \frac{c_w w_l}{c_g^{l-1-i}} w_j (j-l+1) \left(\frac{1}{w_l} + \frac{1}{w_j} \right) \\ &\leq 2\hat{c}_g w_j + \frac{2c_w}{c_g^{l-1-i}} (j-l+1) (w_j + w_l) \end{aligned}$$

Now we can estimate $\sum_{i < l \leq j < r} a_{i,j}$:

$$\sum_{i < l \leq j < r} a_{i,j} < 2\hat{c}_g \sum_{i < l \leq j < r} w_j + 2c_w c_g^{1-l} \sum_{i < l \leq j < r} (j-l+1)c_g^i (w_j + w_l)$$

The first sum is easy

$$\sum_{i < l \leq j < r} w_j = l \sum_{l \leq j < r} w_j \leq l\tilde{\mu}(K_1)$$

For the second sum we get (recall that $M = r - l$)

$$\begin{aligned} \sum_{i < l \leq j < r} (j-l+1)c_g^i (w_j + w_l) &= \sum_{l \leq j < r} (j-l+1)(w_j + w_l) \sum_{i < l} c_g^i \\ &\leq \frac{c_g^l - 1}{c_g - 1} \sum_{l \leq j < r} (j-l+1)(w_j + w_l) \\ &< \frac{c_g^l}{c_g - 1} M[\tilde{\mu}(K_1) + Mw_l] \end{aligned}$$

Therefore,

$$\sum_{i < l \leq j < r} a_{i,j} = O((l + M)\tilde{\mu}(K_1) + M^2 w_l) \quad (15)$$

At last, if $i < l$ and $r \leq j$, we have:

$$\begin{aligned} a_{i,j} &\leq 2w_i w_j \left[\sum_{k=l}^{r-1} \frac{1}{w_k} + \sum_{k=i}^{l-1} \frac{1}{w_k} + \sum_{k=r}^j \frac{1}{w_k} \right] \\ &\leq 2w_i w_j \left[M \left(\frac{1}{w_l} + \frac{1}{w_{r-1}} \right) + \hat{c}_g \left(\frac{1}{w_i} + \frac{1}{w_j} \right) \right] \\ &\leq \frac{2c_w^2}{c_g^{(l-1-i)+(j-r)}} M(w_l + w_{r-1}) + 2\hat{c}_g(w_i + w_j) \end{aligned}$$

We need to bound these two sums:

$$\begin{aligned} \sum_{i < l, r \leq j} (w_i + w_j) &\leq (m-r) \sum_{i < l} w_i + l \sum_{r \leq j} w_j = O((m-r)w_l + lw_{r-1}) \\ \sum_{i < l, r \leq j} \frac{1}{c_g^{(l-1-i)+(j-r)}} &= c_g^{r-l+1} \sum_{i < l} c_g^i \sum_{r \leq j} c_g^{-j} < c_g^{r-l+1} \cdot \frac{c_g^l}{c_g - 1} \cdot c_g^{-m+1} \cdot \frac{c_g^{m-r}}{c_g - 1} = \hat{c}_g^2 \end{aligned}$$

Thus,

$$\sum_{i < l, r \leq j} a_{i,j} = O(M(w_l + w_r) + (m-r)w_l + lw_r) \quad (16)$$

Now we are ready to estimate $\sum_{i < j} a_{i,j}$. We simply put together (8), (10), (14) and its equivalent for $\sum_{r \leq i < j} a_{i,j}$, (15) and analogous bound for $\sum_{l \leq i < r \leq j} a_{i,j}$, and (16).

$$\sum_{0 \leq i < j < m} a_{i,j} = O((l + M + (m-r) + M^2)\tilde{\mu}(K_1))$$

The number $M \leq \lceil D' / (\eta/2) \rceil$ is easy to bound since all slabs in the “middle” section are of width $\eta/2$. So $M^2 \leq \lceil \frac{4D^2n}{c_\eta^2\delta^2} \rceil$. To bound l , resp. $m - r$, we will use Lemma 25: $l \leq \lceil 2Dn/\delta \rceil$. Putting it all together

$$\sum_{0 \leq i < j < m} a_{i,j} = O \left(\left(\frac{4Dn}{\delta} + \frac{4D^2n}{c_\eta^2\delta^2} \right) \tilde{\mu}(K_1) \right) = O \left(\frac{4D^2n}{c_\eta^2\delta^2} \tilde{\mu}(K_1) \right),$$

where the last expression comes from $D \geq \delta$. If c_a is the constant hidden in the O -notation (notice that it does not depend on any of c_η , c_δ , or c_λ), then

$$\sum_{i < j} a_{i,j} \leq \frac{4c_a D^2 n}{c_\eta^2 \delta^2} \tilde{\mu}(K_1) = \frac{c_0}{2\lambda} \tilde{\mu}(K_1),$$

where $c_0 := 8c_a / (c_\eta^2 c_\lambda)$. This is all we needed to finish the proof of this lemma, see (7). \blacksquare

4.3 Proof of the Poincaré Inequality

Assuming that the Poincaré inequality does not hold, with the help of Lemma 17 we find a contradiction within the set K_0 . More precisely, we show that $\mathcal{E}_{K'}(f, f) \geq c_0 \text{Var}_{K_0} f$ holds for every measurable $K' \subseteq K$ which fits in a box of size $[0, \eta] \times [0, \epsilon]^{n-1}$ and the local conductance does not vary much within K' . First we prove that if each of two “nearby” balls has a large intersection with K then their intersection has a large intersection with K too. It will help us to lower-bound $\mathcal{E}_{K'}(f, f)$. This lemma is by Kannan, Lovász, and Simonovits (see [11, Lemma 3.5]).

Lemma 27 *Let $x, y \in \mathbb{R}^n$, $\|x - y\| < \delta/\sqrt{n}$. Then*

$$\text{vol}_n(K \cap (B(x, \delta) \cap B(y, \delta))) \geq \frac{1}{e+1} \min\{\text{vol}_n(K \cap B(x, \delta)), \text{vol}_n(K \cap B(y, \delta))\}$$

Proof :

Without loss of generality we may assume that $x = -y$. Let $C := B(x, \delta) \cap B(y, \delta)$, $C_x := (x - y) + C$, $C_y := (y - x) + C$. By Lemma 21 we know that the function $g(u) = \text{vol}_n(K \cap (u + C))$ is log-concave, therefore for any z , $\log g(0) \geq \frac{\log g(z) + \log g(-z)}{2}$. So,

$$\text{vol}_n(K \cap C)^2 = g(0)^2 \geq g(x - y)g(y - x) = \text{vol}_n(K \cap C_x)\text{vol}_n(K \cap C_y)$$

$$\text{vol}_n(K \cap C) \geq \min\{\text{vol}_n(K \cap C_x), \text{vol}_n(K \cap C_y)\} \tag{17}$$

We want to get a lower bound on $\text{vol}_n(K \cap C_x)$ in terms of $\text{vol}_n(K \cap C)$ and $\text{vol}_n(K \cap B(x, \delta))$. The set $K \cap B(x, \delta)$ can be partitioned into $K \cap C_x$ and $K \cap (B(x, \delta) \setminus C_x)$. We will estimate the volume of the later set in terms of $\text{vol}_n(K \cap C)$. Let $T_x := B(x, \delta) \setminus (C_x \cup C)$. Suppose that by blowing up C (from the origin $\mathbf{0} = \frac{1}{2}(x + y)$) by a factor of β we cover T_x . Let C' be the blown-up C . Then

$$\text{vol}_n(K \cap (B(x, \delta) \setminus C_x)) \leq \text{vol}_n(K \cap C') \leq \beta^n \text{vol}_n(K \cap C),$$

where the second inequality follows from the convexity of K . In other words, if $z' \in K \cap C'$, then there exists $z \in K \cap C$, $\beta z = z'$. Obviously, $z = \frac{1}{\beta} z' \in C$. Since $\mathbf{0}, z' \in K$, by the convexity of K we get $z = \frac{1}{\beta} z' + (1 - \frac{1}{\beta})\mathbf{0} \in K$.

We will prove that if $\beta = 1 + \frac{4}{4n-1}$, then C' covers T_x . Then,

$$\begin{aligned}\text{vol}_n(K \cap C_x) &= \text{vol}_n(K \cap B(x, \delta)) - \text{vol}_n(K \cap (B(x, \delta) \setminus C_x)) \\ &\geq \text{vol}_n(K \cap B(x, \delta)) - \beta^n \text{vol}_n(K \cap C) \\ &\geq \text{vol}_n(K \cap B(x, \delta)) - e \text{vol}_n(K \cap C)\end{aligned}$$

From (17) we get:

$$\begin{aligned}\text{vol}_n(K \cap C) &\geq \min\{\text{vol}_n(K \cap C_x), \text{vol}_n(K \cap C_y)\} \\ &\geq \min\{\text{vol}_n(K \cap B(x, \delta)), \text{vol}_n(K \cap B(y, \delta))\} - e \text{vol}_n(K \cap C),\end{aligned}$$

proving the lemma.

To finish the proof, we need to show that for $\beta = 1 + \frac{4}{4n-1}$ the set C' covers T_x . Let $z \in T_x$. We want to prove that $\frac{1}{\beta}z \in C$. Let $\alpha = 1/\beta$. Clearly, $z \in B(x, \delta)$, therefore we are left to prove $\alpha z \in B(y, \delta)$, or, equivalently, that $\|\alpha z - y\| \leq \delta$. Let $z = \sigma x + w$ where w is orthogonal to x . Then,

$$\|\alpha z - y\|^2 = \|\alpha \sigma x - y\|^2 + \|\alpha w\|^2 = (\alpha + 1)\|x\|^2 + \alpha^2\|w\|^2$$

Since $z \in B(x, \delta)$, we have $\|w\|^2 + (\sigma - 1)^2\|x\|^2 \leq \delta^2$. By the assumption, $2\|x\| = \|x - y\| \leq \delta/\sqrt{n}$. Therefore

$$\begin{aligned}\|\alpha z - y\|^2 &\leq (\alpha \sigma + 1)^2\|x\|^2 + \alpha^2(\delta^2 - (\sigma - 1)^2\|x\|^2) \\ &= (\alpha^2[2\sigma - 1] + \alpha[2\sigma] + 1)\|x\|^2 + \alpha^2\delta^2 \\ &\leq (3\alpha^2 + 4\alpha + 1)\frac{\delta^2}{4n} + \alpha^2\delta^2 \\ &= ((4n + 3)\alpha^2 + 4\alpha + 1)\frac{\delta^2}{4n},\end{aligned}$$

where the last inequality follows from $\sigma \in [0, 2]$. We want to find α as large as possible so that the coefficient of δ^2 becomes at most 1. Solving

$$(4n + 3)\alpha^2 + 4\alpha + 1 \leq 4n$$

gives us $\alpha \in [-1, 1 - \frac{4}{4n+3}]$, therefore if $\alpha = 1 - \frac{4}{4n+3}$, then $\beta = 1/\alpha = 1 + \frac{4}{4n-1}$. ■

At last, we are ready to obtain the final contradiction.

Proof of Theorem 12:

Let $K' \subseteq K$ be a measurable set which fits in a box $[0, \eta] \times [0, \epsilon]^{n-1}$ in some coordinate system. Suppose l varies within K' by a factor at most c_l , in other words for $x, y \in K'$ we have $l(x)/l(y) \leq c_l$. We show that there exists a set I such that $I \subseteq B(x, \delta) \cap K$ for every $x \in K'$ and

$$\text{vol}_n(I) \geq \frac{1}{c_l(e + 1)} \text{vol}_n(B(x_{\max}, \delta) \cap K), \quad (18)$$

where x_{\max} is such that $l(x_{\max}) = \max_{x \in I} l(x)$. Then

$$\begin{aligned}\mathcal{E}_{K'}(f, f) &= \int_{K'} \tilde{\mu}(x) \left[\frac{1}{2 \text{vol}_n(B(x, \delta) \cap K)} \int_{B(x, \delta) \cap K} (f(x) - f(y))^2 dy \right] dx \\ &\geq \frac{1}{2 \text{vol}_n(B(x_{\max}, \delta) \cap K)} \int_{K'} \tilde{\mu}(x) \left[\int_I (f(x) - f(y))^2 dy \right] dx \\ &= \frac{1}{2 \text{vol}_n(B(x_{\max}, \delta) \cap K)} \int_I \int_{K'} \tilde{\mu}(x) (f(x) - f(y))^2 dx dy\end{aligned}$$

It can be easily verified that $\int_{K'} \tilde{\mu}(x)(f(x) - c)^2 dx$ is minimized for $c = \bar{f}_{K'}$. Therefore

$$\begin{aligned} \mathcal{E}_{K'}(f, f) &\geq \frac{1}{2 \operatorname{vol}_n(B(x_{\max}, \delta) \cap K)} \int_I dy \int_{K'} \tilde{\mu}(x)(f(x) - \bar{f}_{K'})^2 dx \\ &= \frac{\operatorname{vol}_n(I)}{2 \operatorname{vol}_n(B(x_{\max}, \delta) \cap K)} \int_{K'} \tilde{\mu}(x)(f(x) - \bar{f}_{K'})^2 dx \\ &\geq \frac{1}{2c_l(e+1)} \operatorname{Var}_{K'} f, \end{aligned}$$

where the last inequality comes from (18).

On the other hand for $K' = K_0$ from Lemma 17, we have $\mathcal{E}_{K'}(f, f) < c_0 \operatorname{Var}_{K'} f$ where c_0 depends on c_η^2 . Also, by the definition of our slabs, the function l_{\max} varies by a factor at most 2 within each slab and at most 4 within K_0 (because K_0 may consist of two adjacent slabs). We would like to say something about the variability of l itself. Let x_{\min} be a point of minimal l over K_0 and let $x_{\min}^{(1)}$ be its first coordinate. By the Lipschitz inequality, see Lemma 22, part 4, we know that $|\ln l_{\max}(x_{\min}^{(1)}) - \ln l(x_{\min})| \leq \frac{\eta}{\delta} \epsilon \sqrt{n-1}$, or equivalently $l_{\max}(x_{\min}^{(1)})/l(x_{\min}) \leq e^{\frac{\eta}{\delta} \epsilon \sqrt{n-1}}$. Recall that we may choose ϵ as small as needed so we may assume that $l_{\max}(x_{\min}^{(1)})/l(x_{\min}) \leq 2$. For any $y, z \in K_0$ we want to bound $l(y)/l(z)$ by a constant.

$$\frac{l(y)}{l(z)} \leq \frac{l_{\max}(y^{(1)})}{l(x_{\min})} \leq \frac{l_{\max}(y^{(1)})}{l_{\max}(x_{\min}^{(1)})/2} \leq 8$$

Therefore we can choose a dimension-independent c_λ so that $c_0 := \frac{8c_a}{c_\eta^2 c_\lambda} \leq \frac{1}{16(e+1)}$ (because $c_l = 8$ in this case), and we obtain a contradiction.

The very last step in this proof is the construction of the set I . Let A and B be the farthest points in K' . If A and B were in the ‘‘opposite corners’’ of K' , by the triangle inequality we get $\|A - B\| \leq \eta + \epsilon \sqrt{n-1} \leq 2\eta$ (since we may pick ϵ as small as we want). We show that the set $I := B(A, \delta') \cap B(B, \delta') \cap K$ for some $\delta' < \delta$ satisfies the two required conditions, namely

- (a) $I \subseteq B(x, \delta) \cap K$ for every $x \in K'$, and
- (b) $\operatorname{vol}_n(I) \geq \frac{1}{c_l(e+1)} \operatorname{vol}_n(B(x_{\max}, \delta) \cap K)$.

Let $\delta' := \delta - \epsilon \sqrt{n-1}$ and $l'(x) := \operatorname{vol}_n(B(x, \delta') \cap K) / \operatorname{vol}_n B(0, \delta')$ (local conductance w.r.t. δ'). The requirement (a) follows directly from the triangle inequality: Let $C \in I$, $x \in K'$, and x' be the projection of x onto the line \overleftrightarrow{AB} such that the first coordinates of x and x' are equal. Then the triangle ABC lies within I , since I is convex. Thus $\|C - x'\| \leq \max\{\|C - A\|, \|C - B\|\} \leq \delta'$. For the distance from C to x this means: $\|C - x\| \leq \|C - x'\| + \|x' - x\| \leq \delta' + \epsilon \sqrt{n-1} = \delta$.

For (b), we will use Lemma 27:

$$\operatorname{vol}_n(I) \geq \frac{1}{e+1} \operatorname{vol}_n(B(x'_{\min}, \delta') \cap K), \quad (19)$$

where x'_{\min} is a point in K' achieving minimal local conductance w.r.t. δ' . We would like to lower-bound the right-hand side of this inequality in terms of $\operatorname{vol}_n(B(x_{\max}, \delta) \cap K)$. To get from x'_{\min} to x_{\max} we will prove that for sufficiently small ϵ also l' varies by a constant factor over K' .

We assume that for any $x, y \in K'$, $l(x)/l(y) \leq c_l$. Our goal is to show that there exists a constant c'_l such that $l'(x)/l'(y) = \text{vol}_n(B(x, \delta') \cap K)/\text{vol}_n(B(y, \delta') \cap K) \leq c'_l$. It is easy to observe that $l'(x) \geq l(x)$ for every $x \in K$. Now we need to upper-bound l' in terms of l . We can choose ϵ small enough so that $\delta' \geq \delta/2^{1/n}$ and therefore $\text{vol}_n B(0, \delta') \geq \text{vol}_n B(0, \delta)/2$. Then

$$l'(x) = \frac{\text{vol}_n(B(x, \delta') \cap K)}{\text{vol}_n B(0, \delta')} \leq \frac{\text{vol}_n(B(x, \delta) \cap K)}{\text{vol}_n B(0, \delta)/2} = 2l(x)$$

For $l'(x)/l'(y)$ this means:

$$\frac{l'(x)}{l'(y)} \geq \frac{l(x)}{2l(y)} \geq \frac{c_l}{2}$$

So $c'_l := c_l/2$. Applying this to (19) we get

$$\text{vol}_n(I) \geq \frac{1}{e+1} \cdot \frac{\text{vol}_n(B(x_{\max}, \delta') \cap K)}{c'_l} \geq \frac{1}{c'_l(e+1)} \cdot \frac{\text{vol}_n(B(x_{\max}, \delta) \cap K)}{2},$$

where the last inequality comes from the assumption $\delta' \geq \delta/2^{1/n}$ and from the observation that for any $x \in K$ if we shrink $B(x, \delta) \cap K$ from x by a factor δ'/δ , we will get a subset of $B(x, \delta') \cap K$. More precisely, let y be any point in $B(x, \delta) \cap K$ and w.l.o.g. we can change the coordinate system so that $x = 0$. Since $\delta'/\delta < 1$, $x \in K$ and K is convex, we get $\frac{\delta'}{\delta}y \in K$. Thus, $\frac{\delta'}{\delta}(B(x, \delta) \cap K) \subseteq K$. Clearly, $\frac{\delta'}{\delta}(B(x, \delta) \cap K) \subseteq B(x, \delta')$. Therefore

$$\text{vol}_n(B(x_{\max}, \delta') \cap K) \geq \left(\frac{\delta'}{\delta}\right)^n \text{vol}_n(B(x_{\max}, \delta) \cap K) \geq \frac{1}{2} \text{vol}_n(B(x_{\max}, \delta) \cap K)$$

We have proven (b), finishing the proof of the Poincaré inequality. ■

5 From Poincaré for Speedy Walk to Uniform Sampling

5.1 Overview

We know the bound on mixing time of the speedy walk from the Poincaré inequality. However, as mentioned in the introduction, the speedy walk is a nice concept but cannot be implemented in real applications. Instead, we need to estimate the mixing time of the Metropolis walk. Assuming that we have a reasonably nice starting distribution, we will prove that the Metropolis walk converges to the stationary distribution quickly (Theorem 34).

The next trick is to obtain a good starting distribution. We assume that $B(0, 1) \subseteq K$ and pick a starting point according to the speedy distribution in the body $K_0 := B(0, 1)$. Then we define a “chain of bodies”, i.e. intersections of K with concentric balls doubling in volume. We will run the Metropolis walk in K_i (where K_i is the intersection of K with the i -th ball), starting at the point in K_{i-1} returned by the previous Metropolis walk. Eventually, we will get a sample point in K with distribution close to the speedy distribution. This is summarized in Algorithm 35 and Theorem 37.

Finally, we want to get from speedy to the uniform distribution. To do this, we will “shrink” K appropriately, obtaining K' . We will sample points in K according to the speedy distribution until we achieve a point in K' . The speedy density at this point will be close to the uniform density and

we output the corresponding point in K (projection of the point in K' onto K). Formalization of this intuition is captured in Algorithm 38 and Theorem 40.

When computing the volume of a convex body, we need to obtain a sample of several (almost) independent points. To make the picture complete we include Algorithm 41 and Theorem 42 dealing with this case.

The algorithms and theorems in this section are only minor modifications of the results in [11, Section 4]. However, the authors of [11] obtained a bound on conductance of the speedy walk and therefore use the so-called M -distance to measure proximity of distributions. We re-state and re-prove their theorems for the Poincaré constant and L_2 -distance.

5.2 More Definitions

We will need some definitions from probability theory. The first definition deals with independence of random variables. In our setting we cannot hope for true independence but as it turns out, “almost” independent random variables (samples) are all we need.

Definition 28 *Two random variables X and Y are ε -independent if for every two measurable sets $A, B \subseteq \Omega$,*

$$|\Pr(X \in A, Y \in B) - \Pr(X \in A)\Pr(Y \in B)| \leq \varepsilon$$

A set of random variables $X_1, \dots, X_k \in \Omega$ is said to be ε -good for a distribution π if

- (a) $d_{tv}(\pi_i, \pi) < \varepsilon$, where π_i is the distribution of X_i ,
- (b) for every $i \neq j$ the variables X_i and X_j are ε -independent.

The next definition deals with expected value conditioned on a non-“exception” event of large probability, the so-called expected value with exception. Unfortunately we are not able to bound the real expected running time of all algorithms presented in this work but we can bound their expected running time with a small exception (authors of [11] conjecture that the theorems remain valid with “ordinary” expectation but can prove them only for “expectation with exception”). Good news is that if the exception is small, we can modify the original algorithm so that it always finishes within prescribed number of steps and it will succeed with high probability. This is more precisely summarized in Note 30.

Definition 29 *We say that a random variable X has expectation at most E with exception ε if there exists $A \subseteq \Omega$ such that $\Pr(A) \geq 1 - \varepsilon$ and the expected value of X conditioned on A is at most E .*

Note 30 *Suppose we proved that an algorithm runs in expected time T_0 with exception ε . This algorithm takes more than $2T_0$ steps with probability at most $\frac{1}{2}(1 - \varepsilon) + \varepsilon$. We can run this algorithm for at most $2T_0$ steps and if it has not finished, we forget its computation and start over. If we repeat this k times, the probability of not succeeding within these $2kT_0$ steps is $1 - (\frac{1+\varepsilon}{2})^k = 1 - 2^{-\Omega(k)}$, assuming $1/2 - \varepsilon = \Omega(1)$. Notice that this modified algorithm takes at most $2kT_0$ steps, as opposed to only having a bound on its expected running time.*

An important parameter in determining the mixing time is the average local conductance, as we will see in Theorem 34.

Definition 31 Average local conductance is defined as $\Lambda := \frac{1}{\text{vol}_n(K)} \int_K l(x) dx$.

5.3 Realization of the Speedy Walk

As mentioned earlier, we are not able to run speedy walk in practice. If we were, after polynomial number of steps t we would be sufficiently close to the stationary distribution $\tilde{\mu}$. This idea is more precisely summarized in the following theorem.

Theorem 32 Let $\delta < c_\delta D/\sqrt{n}$, where c_δ is a dimension-independent constant. Let μ_0 be an initial probability distribution such that $\|\mu_0 - \tilde{\mu}\|_2 < \infty$ and let x_0, x_1, x_2, \dots be the random variables corresponding to the state of the speedy walk on K starting from distribution μ_0 , x_t is the state after t steps. Then

1.

$$\|\mu_t - \tilde{\mu}\|_2 \leq \|\mu_0 - \tilde{\mu}\|_2 e^{-\frac{\lambda t}{2}}$$

2. For $\tau = (\|\mu_0 - \tilde{\mu}\|_2 + \frac{1}{2}) e^{-\frac{\lambda t}{2}}$ the random variables x_0 and x_t are τ -independent.

Proof :

The first claim follows directly from Theorem 6, the Poincaré inequality (Theorem 12), and from $(1 - \lambda/2)^t \leq e^{-\lambda t/2}$.

For the τ -independence, let A, B be any two measurable sets. We want to bound

$$|\Pr(x_0 \in A, x_t \in B) - \Pr(x_0 \in A) \Pr(x_t \in B)| = \mu_0(A) |\Pr(x_t \in B \mid x_0 \in A) - \mu_t(B)| \quad (20)$$

by τ . Let μ'_t be the distribution after t steps starting from distribution μ_0 restricted to the set A . Then (20) equals $\mu_0(A) |\mu'_t(B) - \mu_t(B)| \leq \mu_0(A) (|\mu'_t(B) - \tilde{\mu}(B)| + |\mu_t(B) - \tilde{\mu}(B)|)$. Using the first claim, we will bound $|\mu'_t(B) - \tilde{\mu}(B)|$ and $|\mu_t(B) - \tilde{\mu}(B)|$ in terms of $\|\mu_0 - \tilde{\mu}\|_2$ separately.

Recall that $2d_{tv}(\pi, \tilde{\mu}) = \int_K |\pi(x) - \tilde{\mu}(x)| dx \leq \|\pi - \tilde{\mu}\|_2$ (see (1)), and by definition of d_{tv} we have $|\pi(C) - \tilde{\mu}(C)| \leq d_{tv}(\pi, \tilde{\mu})$ for any measurable set C . Therefore

$$|\mu_t(B) - \tilde{\mu}(B)| \leq \frac{1}{2} \|\mu_t - \tilde{\mu}\|_2 \leq \frac{1}{2} \|\mu_0 - \tilde{\mu}\|_2 e^{-\lambda t/2},$$

where the last inequality comes from the first part of this theorem.

We are left to bound $|\mu'_t(B) - \tilde{\mu}(B)|$. We can use the same argument as before to get $|\mu'_t(B) - \tilde{\mu}(B)| \leq \frac{1}{2} \|\mu'_0 - \tilde{\mu}\|_2 e^{-\lambda t/2}$, where μ'_0 is the initial distribution μ_0 restricted to the set A . Thus, $\mu'_0(x) = \frac{\mu_0(x)}{\mu_0(A)}$. Now we want to express $\|\mu'_0 - \tilde{\mu}\|_2$ in terms of $\|\mu_0 - \tilde{\mu}\|_2$:

$$\|\mu'_0 - \tilde{\mu}\|_2 = \sqrt{\int_K \frac{(\frac{\mu_0(x)}{\mu_0(A)} - \tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx} = \frac{1}{\mu_0(A)} \sqrt{\int_K \frac{(\mu_0(x) - \mu_0(A)\tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx} \quad (21)$$

Since $\mu_0(A) \leq 1$, basic properties of absolute values give us $|\mu_0(x) - \mu_0(A)\tilde{\mu}(x)| \leq |\mu_0(x) - \tilde{\mu}(x)| + |\tilde{\mu}(x) - \mu_0(A)\tilde{\mu}(x)| \leq |\mu_0(x) - \tilde{\mu}(x)| + \tilde{\mu}(x)$. Therefore

$$\begin{aligned} \int_K \frac{(\mu_0(x) - \mu_0(A)\tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx &\leq \int_K \frac{(|\mu_0(x) - \tilde{\mu}(x)| + \tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx \\ &= \int_K \frac{(\mu_0(x) - \tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx + 2 \int_K |\mu_0(x) - \tilde{\mu}(x)| dx + 1 \\ &\leq \|\mu_0 - \tilde{\mu}\|_2 + 2\|\mu_0 - \tilde{\mu}\|_2 + 1 \\ &= (\|\mu_0 - \tilde{\mu}\|_2 + 1)^2 \end{aligned}$$

For (21) this implies

$$\|\mu'_0 - \tilde{\mu}\|_2 \leq \frac{1}{\mu_0(A)} (\|\mu_0 - \tilde{\mu}\|_2 + 1)$$

We bounded all quantities we needed. Collecting all the terms we get

$$\begin{aligned} \mu_0(A) |\Pr(x_t \in B \mid x_0 \in A) - \mu_t(B)| &\leq \mu_0(A) \left(\frac{1}{2} \cdot \frac{1}{\mu_0(A)} (\|\mu_0 - \tilde{\mu}\|_2 + 1) e^{-\lambda t/2} + \frac{1}{2} \|\mu_0 - \tilde{\mu}\|_2 e^{-\lambda t/2} \right) \\ &= \frac{1}{2} e^{-\lambda t/2} (\|\mu_0 - \tilde{\mu}\|_2 + 1 + \mu_0(A) \|\mu_0 - \tilde{\mu}\|_2) \\ &\leq \frac{1}{2} (2\|\mu_0 - \tilde{\mu}\|_2 + 1) e^{-\lambda t/2} = (\|\mu_0 - \tilde{\mu}\|_2 + \frac{1}{2}) e^{-\lambda t/2} \end{aligned}$$

This proves that $|\Pr(x_0 \in A, x_t \in B) - \Pr(x_0 \in A) \Pr(x_t \in B)| \leq \tau$ (see (20)), finishing the proof of τ -independence of x_0 and x_t . \blacksquare

We can simulate the speedy walk using the Metropolis walk which we can easily implement in practice. The previous theorem gives us a bound for the expected number of steps of the speedy walk until it gets sufficiently close to its stationary distribution. However, for the simulation to be time-efficient, we need to be able to bound the number of non-speedy steps, i.e. steps when we stay in the same state. We examine it in the next theorem.

Algorithm 33 (Realization of the Speedy Walk) *Run Metropolis walk from a distribution μ_0 . For given precision $\varepsilon > 0$, stop after $t = \lceil \frac{2}{\lambda} \ln(\frac{5}{2\varepsilon}) \rceil$ proper (i.e. speedy) steps.*

Theorem 34 *Let $\delta < c_\delta D / \sqrt{n}$, where c_δ is a dimension-independent constant. Then*

1. *Let μ_0 be the starting distribution such that $\|\mu_0 - \tilde{\mu}\|_2 \leq 2$, and μ_i be the distribution after i (non necessarily proper) steps of Algorithm 33. Let f be the total number of steps. Then $\|\mu_f - \tilde{\mu}\|_2 \leq \varepsilon$ and the starting point x_0 and the final point x_f are ε -independent.*
2. *If $m_0 := \|\mu_0 - \tilde{\mu}\|_2 < 1$, then the expectation of f with exception m_0^2 is at most $\frac{3t}{\Lambda(1-m_0^2)}$.*

Proof :

First part of this theorem follows directly from Theorem 32 after substituting the expression for t from Algorithm 33.

For the second part, let $V := \{x \in K \mid \frac{\mu_0(x)}{\tilde{\mu}(x)} \geq 3\}$. This set will be our exception, in other words we will show that the expected number of proper and improper steps conditioned on starting the walk outside of V is at most $\frac{3t}{\Lambda(1-m_0^2)}$.

First we want to bound $\mu_0(V)$:

$$\begin{aligned}
m_0^2 &= \int_K \frac{(\mu_0(x) - \tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx \geq \int_V \frac{(\mu_0(x) - \tilde{\mu}(x))^2}{\tilde{\mu}(x)} dx = \\
&= \int_V \frac{\mu_0(x)}{\tilde{\mu}(x)} \mu_0(x) dx - 2 \int_V \mu_0(x) dx + \int_V \tilde{\mu}(x) dx \geq \\
&\geq \int_V 3\mu_0(x) dx - 2\mu_0(V) + \tilde{\mu}(V) \geq 3\mu_0(V) - 2\mu_0(V) = \mu_0(V)
\end{aligned}$$

Therefore $\mu_0(K \setminus V) \geq 1 - m_0^2$. According to Definition 29, let A be the event that we start the walk outside of V (then by the above $\Pr(A) \geq 1 - m_0^2$), and let X be the total number of steps taken by the algorithm. We want to show that $E(X|A) \leq \frac{3t}{\Lambda(1-m_0^2)}$. Let x_i be the state after i proper (speedy) steps, conditioned on the event A , and let μ'_i be the corresponding probability distribution. Clearly, for all $x \notin V$ we have $\mu_0(x) < 3\tilde{\mu}(x)$ and thus

$$\mu'_0(x) = \frac{\mu_0(x)}{\mu_0(K \setminus V)} \leq \frac{3\tilde{\mu}(x)}{1 - m_0^2}$$

By induction on i we will prove that for any $x \in K$

$$\mu'_i(x) \leq \frac{3}{1 - m_0^2} \cdot \tilde{\mu}(x)$$

Let $i > 0$ and suppose the statement to be true for $i - 1$. Then

$$\begin{aligned}
\mu'_i(x) &= \int_K \mu'_{i-1}(y)P(y, x) dy \leq \frac{3}{1 - m_0^2} \int_K \tilde{\mu}(y)P(y, x) dy = \\
&= \frac{3}{1 - m_0^2} \int_K \tilde{\mu}(x)P(x, y) dy = \frac{3}{1 - m_0^2} \tilde{\mu}(x) \int_K P(x, y) dy \leq \frac{3}{1 - m_0^2} \tilde{\mu}(x),
\end{aligned}$$

where P is the transition distribution of the speedy walk on K and the second equality comes from the reversibility of this walk.

Now we are ready to bound the expected number of improper steps between two consecutive proper steps, conditioned on the event A . Clearly, expected number of improper steps before taking a proper move from a state x is $1/l(x)$ (this is the number of improper steps plus the proper move). Thus, conditioned on A the expected number of steps needed to get from x_i to x_{i+1} is

$$\begin{aligned}
\int_K \mu'_i(x) \frac{1}{l(x)} dx &\leq \frac{3}{1 - m_0^2} \int_K \tilde{\mu}(x) \frac{1}{l(x)} dx = \frac{3}{1 - m_0^2} \int_K \left(\frac{l(x)}{\int_K l(y) dy} \right) \frac{1}{l(x)} dx = \\
&= \frac{3}{1 - m_0^2} \cdot \frac{1}{\int_K l(y) dy} \text{vol}_n(K) = \frac{3}{\Lambda(1 - m_0^2)}
\end{aligned}$$

The statement of the theorem follows from the linearity of expectation. ■

5.4 Obtaining a Suitable Starting Distribution

Why cannot we just run Algorithm 33 to obtain a sample point from the speedy distribution? The problem is that we do not have a good starting distribution to pick the initial point for the speedy

walk from. By Theorem 34 we need $\|\mu_0 - \tilde{\mu}\|_2$ to be small enough. In this section we present a method which overcomes this problem.

We assume $B(0, 1) \subseteq K \subseteq B(0, D)$ and we define *intermediate concentric balls* as follows: For $i = 0, \dots, b$ let $B_i := B(0, 2^{i/n})$, where $b = \lceil n \log D \rceil$. In other words, the volume of the balls doubles from i to $i + 1$ and the last ball is guaranteed to cover all of the body K . Further, let $K_i := B_i \cap K$ and let $\tilde{\mu}_i$ be the speedy distribution on K_i (the step-size δ is fixed beforehand). Inductively we will show that we can simulate distributions π_i with L_2 -distance within ε from $\tilde{\mu}_i$. Thus the last distribution π_b is a good approximation of the distribution $\tilde{\mu}_b = \tilde{\mu}$. For convenience we define l_i as the local conductance within K_i , i.e. $l_i(x) := \frac{\text{vol}_n(B(x, \delta) \cap K_i)}{\text{vol}_n B(0, \delta)}$.

Algorithm 35 (Sampling from Speedy Distribution) *Given $\varepsilon < \frac{1}{2}$ do:*

1. Let $\delta < \frac{2}{5\sqrt{n}}$.
2. Let x_0 be chosen according to the speedy distribution $\tilde{\mu}_0$ from K_0 . (This can be done by choosing a point x uniformly from $B_0 = K_0$ and choosing a point y uniformly from $B(x, \delta)$. If $y \in K_0$ then $x_0 := x$, otherwise repeat. This clearly picks x_0 according to $\tilde{\mu}_0$. Also, the number of trial points x is expected to be constant since $l_0(z)$ is lower-bounded by a constant for every $z \in K_0$.)
3. For $i = 1, \dots, b$ do: Run Algorithm 33 on convex body K_i from starting point x_{i-1} . Let x_i be the point returned by this algorithm.
4. Return $y = x_b$.

Before we analyze this algorithm, we state a geometric lemma by Kannan, Lovász, and Simonovits (see [11, Corollary 4.6]) lower-bounding the average local conductance in terms of radius of a ball contained within the convex body K .

Lemma 36 *If $B(0, r) \subseteq K$, then $\Lambda \geq 1 - \frac{\delta\sqrt{n}}{2r}$.*

Theorem 37 *Let π_i be the distribution of the point x_i generated in the i -th iteration of the Algorithm 35. Then $\|\pi_i - \tilde{\mu}_i\|_2 \leq \varepsilon$ for every $i \in \{0, \dots, b\}$.*

The total number of oracle calls with exception $b\varepsilon^2$ is expected to be less than

$$\frac{10bt}{1 - \varepsilon^2} = O\left(n^3 D^2 \ln D \ln \frac{1}{\varepsilon}\right),$$

where $t = \lceil \frac{2}{\lambda} \ln(\frac{5}{2\varepsilon}) \rceil$, and λ is the Poincaré constant of the body K .

Proof :

Not surprisingly, for every i we prove that if $\|\pi_i - \tilde{\mu}_i\|_2 \leq \varepsilon$, then $\|\pi_i - \tilde{\mu}_{i+1}\|_2 \leq 2\varepsilon$. This will allow us to apply Theorem 34 and conclude that $\|\pi_{i+1} - \tilde{\mu}_{i+1}\|_2 \leq \varepsilon$. Since $\pi_0 = \tilde{\mu}_0$, we would have $\|\pi_i - \tilde{\mu}_i\|_2 \leq \varepsilon$ for every i .

To compare $\|\pi_i - \tilde{\mu}_i\|_2$ with $\|\pi_i - \tilde{\mu}_{i+1}\|_2$ we need to be able to express $\tilde{\mu}_i(x)$ in terms of $\tilde{\mu}_{i+1}(x)$.

$$\tilde{\mu}_i(x) = \frac{l_i(x)}{\int_{K_i} l_i(y) dy} \leq \frac{l_{i+1}(x)}{\int_{K_i} l_i(y) dy} = \frac{\int_{K_{i+1}} l_{i+1}(y) dy}{\int_{K_i} l_i(y) dy} \tilde{\mu}_{i+1}(x)$$

We will bound the ratio $\int_{K_{i+1}} l_{i+1} / \int_{K_i} l_i$ using Λ_i , the average local conductance of the i -th body K_i . Since $l_{i+1}(y) \leq 1$, clearly $\int_{K_{i+1}} l_{i+1} \leq \text{vol}_n(K_{i+1})$. On the other hand $\int_{K_i} l_i = \Lambda_i \text{vol}_n(K_i)$:

$$\frac{\tilde{\mu}_i(x)}{\tilde{\mu}_{i+1}(x)} \leq \frac{\int_{K_{i+1}} l_{i+1}(y) dy}{\int_{K_i} l_i(y) dy} \leq \frac{\text{vol}_n(K_{i+1})}{\Lambda_i \text{vol}_n(K_i)} \leq \frac{2}{\Lambda_i} \quad (22)$$

Therefore

$$\begin{aligned} \|\pi_i - \tilde{\mu}_{i+1}\|_2^2 &= \int_{x \in K} \frac{\pi_i(x)^2}{\tilde{\mu}_{i+1}(x)} dx - 1 \leq \int_{x \in K} \frac{\pi_i(x)^2}{\frac{\Lambda_i}{2} \tilde{\mu}_i(x)} dx - 1 = \\ &= \frac{2}{\Lambda_i} \|\pi_i - \tilde{\mu}_i\|_2^2 + \frac{2}{\Lambda_i} - 1 \leq \frac{2}{\Lambda_i} \cdot \varepsilon^2 + \frac{2}{\Lambda_i} - 1 \end{aligned}$$

Since $B(0, 1) \subseteq K_i$ for every i , we may use Lemma 36 to bound $\Lambda_i \geq 4/5$. Thus we get $\|\pi_i - \tilde{\mu}_{i+1}\|_2^2 \leq 2(\varepsilon^2 + 1)/\Lambda_i - 1 \leq 5 - 1 = 4$, proving $\|\pi_i - \tilde{\mu}_{i+1}\|_2 \leq 2$.

For the second part of the theorem, to bound the number of oracle calls, we will follow the proof of Theorem 34, part 2. More precisely, we will show that for every i the expected number of oracle calls with exception ε^2 is less than $10t_i/(1 - \varepsilon^2)$, where t_i is the number of speedy steps of Algorithm 33 run in the body K_i . Clearly, the diameter D_i of the body K_i increases with i , thus by Theorem 12 the λ_i decrease with i and therefore $t_0 \leq t_1 \leq \dots \leq t_b = t$. So by linearity of expectation we expect the total number of steps to be less than $10bt/(1 - \varepsilon^2)$ with exception $b\varepsilon^2$.

In this case, let $V_i := \{x \in K_i \mid \frac{\pi_i(x)}{\tilde{\mu}_i(x)} \geq 3\}$ be our exception set in the i -th body. (Notice that this set is slightly different from the one in Theorem 34, part 2. There V_i would be the set of all $x \in K_i$ such that $\pi_i(x) \geq 3\tilde{\mu}_{i+1}(x)$.)

Mimicking the proof of Theorem 34, part 2, for this new exception set we get that $\pi_i(V_i) \leq \|\pi_i - \tilde{\mu}_i\|_2^2 \leq \varepsilon^2$. Similarly, let A be the event that we start the i -th walk outside of V_i and let X be the number of steps taken by the i -th iteration of the Algorithm 33. We can bound $E(X|A)$ as before: Let $y_{i,j}$ be the state after j speedy steps starting the walk from $x_i \notin V$, i.e. $y_{i,0} := x_i$, and let $\pi'_{i,j}$ be the corresponding probability distribution. Then for $y \notin V$

$$\pi'_{i,j}(y) = \frac{\pi_i(y)}{\pi_i(K \setminus V)} \leq \frac{3\tilde{\mu}_i(y)}{1 - \varepsilon^2} \leq \frac{7.5\tilde{\mu}_{i+1}(y)}{1 - \varepsilon^2},$$

where the last inequality comes from (22). By induction on j we get $\pi'_{i,j}(y) \leq \frac{7.5}{1 - \varepsilon^2} \tilde{\mu}_{i+1}(y)$ for any $y \in K_i$. Now we can bound the expected number of steps of the i -th walk, conditioned on the event A : $E(X|A) \leq \frac{7.5t_i}{\Lambda_i(1 - \varepsilon^2)} < \frac{10t_i}{1 - \varepsilon^2}$. ■

5.5 From Speedy Distribution to the Uniform Distribution

Algorithm 38 (Sampling from Uniform Distribution) *Let \mathcal{A} be an algorithm generating samples from K . Using \mathcal{A} generate samples y_1, y_2, \dots until a point y_k is obtained such that $z := \frac{2n}{2n-1}y_k \in K$. Return z .*

We will use this well-known geometric statement for the analysis of this algorithm.

Lemma 39 *Let H be a half-space not containing the point y . If the distance t of H from y satisfies $\delta/\sqrt{n} < t$, then*

$$\frac{\text{vol}_n(H \cap B(y, \delta))}{\text{vol}_n B(0, \delta)} < e^{-nt^2/(2\delta^2)}$$

Theorem 40 *Let $\delta \leq \frac{1}{\sqrt{8n \ln(1/\varepsilon)}}$ and $\varepsilon < 1/10$. If π is the distribution associated with algorithm \mathcal{A} such that $d_{tv}(\pi, \tilde{\mu}) \leq \varepsilon$, and π' is the distribution of Algorithm 38, then $d_{tv}(\pi', \mu) < 11\varepsilon$. Further, the expected number k of samples y_i needed in Algorithm 38 is less than 5.*

Proof :

Let $\alpha = 1 - \frac{1}{2n}$ and $K' := \alpha K$. We search for the first y_i in K' . Let Λ' be the average local conductance of K' , i.e. $\Lambda' := \frac{1}{\text{vol}_n(K')} \int_{K'} l(y) dy$. We will show that $\Lambda' > 1 - \varepsilon$, allowing us to conclude that $\pi(K') > 1/5$, as well as $d_{tv}(\pi', \mu) < 11\varepsilon$.

Let us assume $\Lambda' > 1 - \varepsilon$. For the total variation distance, we want to prove that for any measurable set $S \subseteq K$, the difference $|\pi'(S) - \mu(S)| < 11\varepsilon$. Since $\pi'(z) = \pi(\alpha z)/\pi(K')$, we get

$$\pi'(S) - \mu(S) = \frac{\pi(\alpha S)}{\pi(K')} - \mu(S) \leq \frac{\tilde{\mu}(\alpha S) + \varepsilon}{\tilde{\mu}(K') - \varepsilon} - \mu(S),$$

where the last inequality comes from the assumption about the total variation distance of $\tilde{\mu}$ and π .

$$\tilde{\mu}(\alpha S) = \frac{\int_{\alpha S} l(y) dy}{\int_K l(y) dy} \leq \frac{\text{vol}_n(\alpha S)}{\int_K l(y) dy}$$

Similarly,

$$\tilde{\mu}(K') = \Lambda' \frac{\text{vol}_n(K')}{\int_K l(y) dy} \geq (1 - \varepsilon) \frac{\text{vol}_n(K')}{\int_K l(y) dy} \quad (23)$$

By definition of uniform distribution, $\mu(x) = 1/\text{vol}_n(K)$. Thus $\mu(S) = \frac{\text{vol}_n(S)}{\text{vol}_n(K)} = \frac{\text{vol}_n(\alpha S)}{\text{vol}_n(K')}$. For $\pi'(S) - \mu(S)$ we get

$$\pi'(S) - \mu(S) \leq \frac{\text{vol}_n(\alpha S) + \varepsilon \int_K l(y) dy}{(1 - \varepsilon)\text{vol}_n(K') - \varepsilon \int_K l(y) dy} - \frac{\text{vol}_n(\alpha S)}{\text{vol}_n(K')}$$

Since $\alpha^{-n} \leq e^{n/(2n-1)} \leq e$, we can bound $\int_K l(y) dy \leq \text{vol}_n(K) = (\frac{1}{\alpha})^n \text{vol}_n(K') \leq e \text{vol}_n(K')$. Then

$$\begin{aligned} \pi'(S) - \mu(S) &\leq \frac{\text{vol}_n(\alpha S) + \varepsilon e \text{vol}_n(K') - (1 - \varepsilon - \varepsilon e)\text{vol}_n(\alpha S)}{(1 - \varepsilon - \varepsilon e)\text{vol}_n(K')} \\ &= \frac{\varepsilon e}{1 - \varepsilon - \varepsilon e} + \frac{\varepsilon(1 + e)}{1 - \varepsilon - \varepsilon e} \cdot \frac{\text{vol}_n(\alpha S)}{\text{vol}_n(K')} \\ &\leq \frac{2\varepsilon e + \varepsilon}{1 - \varepsilon - \varepsilon e} \leq \frac{2e + 1}{1 - \frac{1}{10}(1 + e)} \varepsilon < 11\varepsilon, \end{aligned}$$

where in the second inequality we used $\alpha S \subseteq K'$, and the third inequality holds for sufficiently small ε (i.e. $\varepsilon < 1/10$). Thus we get $\pi'(S) - \mu(S) < 11\varepsilon$, proving $d_{tv}(\pi', \mu) < 11\varepsilon$.

Now we can bound $\tilde{\mu}(K')$ using (23).

$$\tilde{\mu}(K') \geq (1 - \varepsilon) \frac{\alpha^n \text{vol}_n(K)}{\int_K l(y) dy} = (1 - \varepsilon) \frac{\alpha^n}{\Lambda} \geq \frac{1 - \varepsilon}{e\Lambda} > \frac{1}{5} + \varepsilon,$$

where the last inequality comes from $\Lambda \leq 1$ and $\varepsilon < 1/10$. By the assumption on the total variation distance of $\tilde{\mu}$ and π , we know that $\pi(K') > 1/5$, allowing us to conclude that the expected number of y_i computed by the algorithm is less than 5.

At last, we need to prove the bound on Λ' , i.e. $\Lambda' > 1 - \varepsilon$. We will show that $\int_{K'} (1 - l(x)) dx < \varepsilon \text{vol}_n(K')$ from where the Λ' bound follows directly. Since $1 - l(x) = \frac{\text{vol}_n(B(x, \delta) \cap K^c)}{\text{vol}_n B(0, \delta)}$, we have

$$\int_{K'} (1 - l(x)) dx = \frac{1}{\text{vol}_n B(0, \delta)} \text{vol}_{2n} \{(x, y) \mid x \in K', y \notin K, \|x - y\| \leq \delta\}$$

Therefore

$$\int_{K'} (1 - l(x)) dx = \int_{t>0} \frac{1}{\text{vol}_n B(0, \delta)} \int_{(1+t+dt)K \setminus (1+t)K} \text{vol}_n(B(y, \delta) \cap K') dy dt$$

Let $y \notin (1+t)K$. Using Lemma 39 we can bound $\text{vol}_n(B(y, \delta) \cap K')$: Let ϱ be the hyperplane separating y from K tangent to $(1+t)K$. Let d_ϱ be the distance from the origin to the corresponding hyperplane tangent to K . The perpendicular distance from y to the corresponding hyperplane tangent to K' is at least $(1+t-\alpha)d_\varrho$. Since $B(0, 1) \subseteq K$, we have $d_\varrho \geq 1$ and the distance is at least $1+t-\alpha = t + \frac{1}{2n}$. Now we can use Lemma 39:

$$\begin{aligned} \text{vol}_n(B(y, \delta) \cap K') &\leq e^{-n(t+\frac{1}{2n})^2/(2\delta^2)} \text{vol}_n B(0, \delta) \\ &= e^{-1/(8n\delta^2)} e^{-t/(2\delta^2)} e^{-nt^2/(2\delta^2)} \text{vol}_n B(0, \delta) \leq e^{-1/(8n\delta^2)} e^{-t/(2\delta^2)} \text{vol}_n B(0, \delta) \end{aligned}$$

For $\int_{K'} (1 - l(x)) dx$ this means:

$$\int_{K'} (1 - l(x)) dx \leq \int_{t>0} \int_{(1+t+dt)K \setminus (1+t)K} e^{-1/(8n\delta^2)} e^{-t/(2\delta^2)} dy dt$$

Since $(1+t)K \subseteq (1+t+dt)K$ (because K contains the origin), we can estimate the volume of $(1+t+dt)K \setminus (1+t)K$:

$$\begin{aligned} \text{vol}_n((1+t+dt)K \setminus (1+t)K) &= \text{vol}_n((1+t+dt)K) - \text{vol}_n((1+t)K) \\ &= [(1+t+dt)^n - (1+t)^n] \text{vol}_n(K) \\ &= n(1+t)^{n-1} \text{vol}_n(K) dt \leq ne^{nt} \text{vol}_n(K) dt \end{aligned}$$

Thus,

$$\int_{K'} (1 - l(x)) dx \leq \text{vol}_n(K) \int_{t>0} e^{-1/(8n\delta^2)} e^{-t/(2\delta^2)} ne^{nt} dt \leq e^{-1/(8n\delta^2)} \text{vol}_n(K) < \varepsilon \text{vol}_n(K)$$

For the second inequality one needs to compute one-dimensional integral $\int_{t>0} e^{-t/(2\delta^2)+nt} dt \leq \frac{1}{n}$. The last inequality follows from substituting for δ . This concludes the proof of this theorem. \blacksquare

Algorithm 41 (Sampling N almost independent points from the uniform distribution.)

Given is $\varepsilon > 0$, $N \in \mathbb{N}$, and K , a convex body such that $B(0, 1) \subseteq K \subseteq B(0, D)$.

1. Let x_1 be the point returned by Algorithm 35.
2. Let $i := 2$ and $j := 0$.
3. While $j < N$ do: Run Algorithm 33 from x_{i-1} and obtain x_i . If $\frac{2n}{2n-1}x_i \in K$, then $j := j + 1$, and $y_j := x_i$.
4. Return y_1, \dots, y_N .

Theorem 42 Algorithm 41 uses $O^*(n^3D^2 + Nn^2D^2)$ oracle calls. If π_j is the distribution of y_j , then π_j is ε -close to the uniform distribution μ in total variation distance, and for every $i \neq j$ the points y_i and y_j are ε -independent.

Proof :

Let σ_i be the distribution of x_i . Then by Theorem 37 the distribution σ_1 is close to $\tilde{\mu}$ in the L_2 -distance: $\|\sigma_1 - \tilde{\mu}\|_2 \leq \varepsilon$. Thus we have a good starting distribution and we may apply Theorem 34 to get $\|\sigma_i - \tilde{\mu}\|_2 \leq \varepsilon$ for every i . By the same theorem the points x_1, x_2, \dots are pairwise ε -independent since for every i and $j > i$ we did get from x_i to x_j using more than t steps of the speedy walk. Recall that $d_{tv}(\sigma_i, \tilde{\mu}) \leq \frac{1}{2}\|\sigma_i - \tilde{\mu}\|_2$ (see (1)). Therefore Theorem 40 gives us $d_{tv}(\pi_j, \mu) < 11/2\varepsilon$.

The expected number of oracle calls in step 1 is $O^*(n^3D^2)$, see Theorem 37. For generating x_i for $i > 1$ we need $O^*(\frac{1}{\lambda}) = O^*(n^2D^2)$ oracle calls, see Theorem 34 and Theorem 12. By Theorem 37 the total expected number of x_i generated by the algorithm is less than $2N$. By Wald's equality this gives us total $O^*(n^3D^2 + Nn^2D^2)$ oracle calls. ■

6 Acknowledgements

I would like to thank my thesis advisor Eric Vigoda for many hours of discussions and numerous e-mails, Mark Jerrum for useful insights and hints (communicated through Eric), Daniel Štefankovič for all his help and support, and Laci Lovász for explaining the main ideas intuitively.

References

- [1] Aldous, Fill
- [2] D. Applegate, and R. Kannan, *Sampling and integration of near log-concave functions*. Proc. 23rd ACM STOC (1990), 156–163.
- [3] Gardner, R. J. *The Brunn-Minkowski inequality*. Bull. Amer. Math. Soc. (N.S.) 39 (2002), no. 3, 355–405.
- [4] M. Dyer, and A. Frieze, *Computing the volume of convex bodies: a case where randomness provably helps*. Probabilistic combinatorics and its applications (San Francisco, CA, 1991), 123–169.

- [5] M. Dyer, A. Frieze, and R. Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*. Proc. of the 21st Annual ACM Symposium on Theory of Computing (1989), 375–381.
- [6] H. G. Eggleston, *Convexity*, Cambridge University Press, New York, 1958.
- [7] A. Frieze, and R. Kannan, *Log-Sobolev inequalities and sampling from log-concave distributions*. Ann. Appl. Probab. 9 (1999), no. 1, 14–26.
- [8] M. Jerrum, unpublished manuscript
- [9] Jerrum, Sinclair
- [10] Jerrum, Son
- [11] R. Kannan, L. Lovász, and M. Simonovits, *Random walks and an $O^*(n^5)$ volume algorithm for convex bodies*. Random Structures Algorithms 11 (1997), no. 1, 1–50.
- [12] L. Lovász, *How to compute the volume?*. Jber. d. Dt. Math-Verein, Jubiläumstagung 1990, B.G. Teubner, Stuttgart, 138–151.
- [13] L. Lovász, and M. Simonovits, *The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume*. Proc. of the 31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990), 346–354.
- [14] L. Lovász, and M. Simonovits, *On the randomized complexity of volume and diameter*. Proc. 33rd IEEE FOCS (1992), 482–491.
- [15] L. Lovász, and M. Simonovits, *Random walks in a convex body and an improved volume algorithm*. Random Structures Algorithms 4 (1993), no. 4, 359–412.
- [16] M. Mihail, *Conductance and convergence of Markov chains - A combinatorial treatment of expanders*. Proc. of the 30th Annual Symposium on Foundations of Computer Science (1989), 526–531.