Combinatorics, Probability and Computing (2005) 00, 000–000. © 2005 Cambridge University Press
DOI: 10.1017/S0000000000000000 Printed in the United Kingdom

Blocking conductance and mixing in random walks

R. KANNAN[1] †, L. LOVASZ[2] and R. MONTENEGRO[3] ‡

[1] Department of Computer Science, Yale University, New Haven, CT 06520
[2] Microsoft Research, One Microsoft Way, Redmond, WA 98052
[3] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160
(email: kannan@cs.yale.edu, lovasz@microsoft.com and monteneg@math.gatech.edu)

The notion of conductance introduced by Jerrum and Sinclair [8] has been widely used to prove rapid mixing of Markov Chains. Here we introduce a bound that extends this in two directions. First, instead of measuring the conductance of the worst subset of states, we bound the mixing time by a formula that can be thought of as a weighted average of the Jerrum-Sinclair bound (where the average is taken over subsets of states with different sizes). Furthermore, instead of just the conductance, which in graph theory terms measures edge expansion, we also take into account node expansion. Our bound is related to the logarithmic Sobolev inequalities, but it appears to be more flexible and easier to compute.

In the case of random walks in convex bodies, we show that this new bound is better than the known bounds for the worst case. This saves a factor of $O(n)$ in the mixing time bound, which is incurred in all proofs as a "penalty" for a "bad start". We show that in a convex body in $\mathbb{R}^n$, with diameter D, random walk with steps in a ball with radius δ mixes in $O^*(nD^2/\delta^2)$ time (if idle steps at the boundary are not counted). This gives an $O^*(n^3)$ sampling algorithm after appropriate preprocessing, improving the previous bound of $O^*(n^4)$.

The application of the general conductance bound in the geometric setting depends on an improved isoperimetric inequality for convex bodies.

1. Introduction

One of the basic algorithmic problems is *sampling*: choosing a random element of a given set, from a given distribution π. Sampling algorithms are applied in statistics, simulation, approximate combinatorial enumeration, volume computation, integration, optimization, and many other areas.

Perhaps the most important method in sampling algorithms is the use of rapidly mixing Markov chains. The general idea is to construct an ergodic Markov chain whose state

† Research partially supported by NSF grant CCR-0310805.
‡ Research partially supported by VIGRE grants at Yale University and the Georgia Institute of Technology.

space V is the set we want to sample from, whose stationary (steady state) distribution is the given distribution π, and whose transitions can be easily simulated. One then has to show that the chain approaches the steady state after a reasonably small number of steps (typically, we want polynomial in the "input size" of the problem). Of these, the last part, which amounts to estimating the "mixing time", is usually the hardest.

Let us define the mixing time. Roughly speaking, this is the smallest T such that the distribution σ^T of the state after T steps will be close to stationary. There is a small technical difficulty: it may be that the chain is periodic (an example is the random walk on a bipartite graph), in which case the distribution of σ^T does not even tend to the stationary. For the purposes of sampling, this periodicity can be eliminated quite easily by choosing the number of steps T at random from some simple distribution: Poisson distribution (continuous time walk), binomial distribution (lazy walk) etc. For the purposes of this paper, we choose T uniformly from the set $\{0, 1, \ldots, t-1\}$ for a given $t > 0$. (This is not the best in practice, but it is best suited for our analysis.) Stopping after a random number T of steps, we get a state from the distribution

$$\rho^t = \frac{1}{t}(\sigma^0 + \cdots + \sigma^{t-1}) \tag{1.1}$$

The other issue to clarify is what we mean by "close". For our purposes, it will be sufficient to require that the total variation distance $d(\sigma^T, \pi) < 1/4$ (in other words, $|\sigma^T(A) - \pi(A)| < 1/4$ for every set $A \subseteq V$). We denote by $\mathcal{T}$ the smallest t for which this is true for every starting node. General results [1] imply that in $k\mathcal{T}$ steps, the total variation distance can be reduced to 2^{-k}.

(Our results would remain valid in the "stopping rule" framework for mixing time (see [2, 17]), and in fact some of the formulas would turn out simpler; but we don't want to go into the measure theoretic technicalities needed in that framework.)

The most common tool used in the proof of rapid mixing is the notion of *conductance* proposed by Sinclair and Jerrum [8] which they describe in the setting of finite-state chains. We recall their notion in that setting now. For a set S of states, we define the *ergodic flow* from S as $Q(S) = \sum_{i \in S} \sum_{j \notin S} \pi_i p_{ij}$; this is the long-run fraction of moves from S to its complement S^c. The conductance of a set S of states is defined by

$$\Phi(S) = \frac{Q(S)}{\pi(S)(1 - \pi(S))}$$

(the denominator is the long-run fraction of moves from S to its complement S^c in a sequence of independent samples from π). The conductance Φ of the chain is the minimum conductance of nonempty proper subsets of V.

We denote by π_0 the minimum stationary probability of any state. In a typical discrete sampling problem, π_0 is exponentially small.

The conductance is related to the mixing time through the following fundamental inequality:

Theorem 1.1 (Jerrum and Sinclair).

$$\frac{1}{\Phi} \lesssim \mathcal{T} \lesssim \ln(1/\pi_0)\frac{1}{\Phi^2}$$

(Here $a \lesssim b$ means that $a \leq cb$ for some absolute constant $c > 0$.) This bound has been used very successfully both in combinatorial chains (e.g. to sample from perfect matchings) as well as for random walks in convex bodies. Typically, the factor of $\ln(1/\pi_0)$ is polynomial in the input size, and then Theorem 1.1 says that the polynomiality of the mixing time $\mathcal{T}$ is equivalent to the polynomiality of $1/\Phi$.

However, if we are interested in a finer analysis of the mixing time, the factor $\ln(1/\pi_0)$ is quite substantial. This factor, sometimes called the "start penalty", accounts for getting stuck in a small set close to the starting state. One way of saving this factor is not to start at a fixed point, but at a random point from a distribution not concentrated on or near a single point. (It is not trivial that we can generate such a random starting point, but sometimes it is possible by bootstrapping or by re-using some previously generated points.) Such an approach was taken, e.g., in [12]; however, the slight dependencies between different parts of the algorithm, introduced by this bootstrapping and recycling, substantially increases the tediousness of the analysis.

Another possibility is to make use of the fact that small sets often have larger boundary, and therefore the walk spreads faster at the beginning than predicted by conductance alone. One way to formalize this analytically is through the use of *logarithmic Sobolev inequalities*, proposed by Diaconis and Saloff-Coste [4]. However, bounding the log-Sobolev constant seems difficult. In the context of rapidly mixing Markov Chains used in counting problems and volume computation, this has been successfully applied only to the "king's moves" [7] (which for other reasons is not efficient) and to the basis-exchange walk for balanced matroids [9], although decomposition methods hold some hope for future applications [10].

In [14] Lovász and Kannan introduced the method of *average conductance*. If we define

$$\phi(x) = \min\{\Phi(S) : \ \pi(S) = x\},$$

then they proved

Theorem 1.2.

$$\mathcal{T} \lesssim \int_{\pi_0}^{1/2} \frac{1}{x\phi(x)^2}.$$

It was sketched how this theorem can be applied to get rid of the "start penalty" in the case of the ball walk in a convex body.

There are other cases, where one would like to take into account that sets of different sizes have different conductance. For example, small sets (of stationary measure at most ε) may have very bad conductance, but one can ignore them: the walk quickly gets closer to the stationary distribution than ε, and the convergence only slows down after that. This idea was exploited in [15], using the notion of "μ-conductance".

Another situation where the Jerrum–Sinclair bound is too pessimistic is the "dumbbell graph": two cliques on $n/2$ nodes joined by an edge. The conductance of this graph is $\Theta(1/n^2)$, and the mixing time $\Theta(n^2)$, less than the square root of the Jerrum–Sinclair upper bound. An obvious heuristic explanation is that there is only one bad cut, and a few small cuts nearby, but most of the cuts have conductance $\Theta(1)$.

A further improvement can be had by examining how large a neighborhood the Markov chain expands into, i.e., the node-expansion. In the extreme case of the complete graph K_n, a single step from any vertex can reach any other vertex and so the mixing time is faster than predicted by conductance alone.

In this paper, we develop a new method for estimating the mixing time, taking into account sets of various sizes, and also both edge-expansion and node-expansion. We note that Morris and Peres [21] have developed a related method known as Evolving Sets in which they bound the stronger relative pointwise distance. Results by these two methods are closely related, although our results are stronger when it comes to mixing time [19].

The main motivation for our study was a problem in connection with the problem of generating a uniformly distributed random point in a convex body K in $\mathbb{R}^n$. One of the most efficient methods for this is to do a random walk in the convex body, stepping from a current point v to a uniformly distributed point in the ball $v + \delta B$ with a given radius δ about v (if this point is not in K, then a new point is generated). This is called the *ball walk with δ-steps*. For a large enough integer T, stopping the walk after T steps, we get an approximately uniform point in K. The main problem is to determine how large T must be.

We try to summarize some previous results on this problem without going into too much detail. Let us assume that K contains the unit ball B, and has diameter D. The conductance of the ball walk in K is $O(\delta/(\sqrt{n}D))$, but there is a "start penalty", and the best previous bound on the mixing time [12] was $O^*(n^2D^2/\delta^2)$. With appropriate preprocessing, one can achieve $D = O^*(\sqrt{n})$, and the best choice of δ is $\Theta^*(1/\sqrt{n})$ (for larger δ, the idle steps at the boundary become dominant), and so one gets a bound of $O^*(n^4)$ on the mixing time.

We apply the result to the ball walk in a convex body, and show that in this case it effectively saves the "start penalty": it is better by a factor of $O^*(n)$ than the above mentioned bound. This application depends on a new geometric isoperimetric inequality, showing that the conductance of small sets is actually better than the worst case bound.

Such a saving is also possible for certain combinatorial chains, as a recent work of Benjamini and Mossel [3] shows.

In the conference version of this paper [14] a somewhat weaker result was stated, and for finite Markov chains only. In this paper, we state and prove our main result in the setting of continuous state Markov chains, which not only means greater generality, but allows us to get stronger results, and give simpler arguments. For our main application to random walks in convex bodies, this general setting is needed. In section 3.5 we indicate how to derive from these similar bounds for finite Markov chains. The derivation is straightforward but quite tedious, so we don't give the details; the interested reader can see [18] for a related argument.

We only consider time reversible chains, which are the most common in applications; some of our techniques extend to non-reversible chains, but the elaboration of the details is left for further work.

In section 2, we introduce the machinery from the theory of continuous state Markov chains. Section 3 contains the statement of our main result and its corollaries, in the setting of continuous state (atom-free) Markov chains. In section 4, we prove an isoperi-

metric inequality for convex bodies, and combine it with our mixing results to give an improved (and in a sense optimal) bound on the mixing time of the ball walk in convex bodies. Sections 5 and 6 contain the proofs of our results.

2. Continuous state Markov chains

Recall the definition of a continuous state Markov chain. Let $(K, \mathcal{A})$ be a σ-algebra. For every $u \in K$, let P_u be a probability measure on $(K, \mathcal{A})$, and assume that for every $S \in \mathcal{A}$, the value $P_u(S)$ is a measurable function of u ($P_u(S)$ denotes the 1-step transition probability density out of u). Then the triple $(K, \mathcal{A}, \{P_u : u \in K\})$ is a *Markov chain*.

For every measure μ on $(K, \mathcal{A})$, we define another measure $P\mu$ by

$$P\mu(S) = \int_K P_x(S)\, d\mu(x).$$

We have to make some assumptions about the Markov chains we consider. We assume that the Markov chain has a *stationary distribution* π. This distribution satisfies

$$P\pi(S) = \pi(S)$$

for every $S \in \mathcal{A}$.

We assume that the chain is *ergodic*, i.e., it escapes from every set $S \in \mathcal{A}$, $0 < \pi(S) < 1$ with positive probability:

$$\int_S P_x(K \setminus S)\, d\pi > 0.$$

We assume that the Markov chain is *time-reversible*, which means that

$$dP_y(x)\, d\pi(y) = dP_x(y)\, d\pi(x).$$

In an integrated form this means that for any two sets $X, Y \in \mathcal{A}$

$$\int_X P_x(Y)\, d\pi(x) = \int_Y P_y(X)\, d\pi(y).$$

We also assume that the probability space $(K, \mathcal{A}, \pi)$ is *atom-free*, i.e., every $X \in \mathcal{A}$ with $\pi(X) > 0$ has a measurable subset $Y \subset X$, $Y \in \mathcal{A}$ with $0 < \pi(Y) < \pi(X)$. (This is not an essential assumption but it is convenient; we'll see in section 3.5 how atoms can be replaced by atom-free sets without changing mixing properties.) It then follows that X has a measurable subset Y with $\pi(Y) = t$ for every $0 < t < \pi(X)$.

2.1. Conductance and other expansion measures

The *ergodic flow* Q is defined for $X, Y \in \mathcal{A}$ by

$$Q(X, Y) = \int_X P_x(Y)\, d\pi(x).$$

By the definition of π, we have $Q(S, K) = Q(K, S) = \pi(S)$; this implies that

$$Q(X, Y) \le \min\{\pi(X), \pi(Y)\}.$$

Also note that by time reversibility, we have $Q(X,Y) = Q(Y,X)$ (even in the non-reversibile case $Q(X,Y) = \pi(Y) - Q(Y,Y) = \pi(Y) - (\pi(Y) - Q(Y,X)) = Q(Y,X)$). Ergodicity means that $Q(S, K \setminus S) > 0$ for all $S \in \mathcal{A}$ with $0 < \pi(S) < 1$.

We define the *conductance function* as follows. For $t \in (0,1)$, let

$$\Psi(t) = \min_{\substack{S \subset V, \\ \pi(S)=t}} Q(S).$$

The conductance function considered in the introduction is

$$\Phi(t) = \frac{\Psi(t)}{t(1-t)};$$

the division by $t(1-t)$ will not be convenient in the more general setting.

It is clear that

$$\Psi(x) = \Psi(1-x) \tag{2.1}$$

and

$$\Psi(x) \le \min\{x, 1-x\} \le 2x(1-x) \le 1/2. \tag{2.2}$$

The conductance function has the Lipschitz property:

Lemma 2.1. *For every $0 < x, y < 1$,*

$$|\Psi(y) - \Psi(x)| \le |y - x|.$$

Proof. We may assume that $x < y$. By (2.1), it is enough to prove that

$$\Psi(y) - \Psi(x) \le y - x.$$

Let $S \subseteq K$ satisfy $\pi(S) = x$ and $Q(S) = \Psi(x)$. Let S' be any set with $S \subseteq S' \subseteq K$ with $\pi(S') = y$. Then we have

$$\begin{aligned} \Psi(y) \le Q(S') &= Q(S, K \setminus S') + Q(S' \setminus S, K \setminus S') \\ &\le Q(S) + Q(S' \setminus S, K) \\ &= \Psi(x) + y - x. \end{aligned}$$

□

Let A be a measurable subset of K. For every $0 \le x \le \pi(A)$, let

$$\Psi(x, K \setminus A) \quad = \quad \inf\{Q(C, K \setminus A) : \ C \subseteq A, \pi(C) = x\}.$$

It will be convenient to extend the definition for $x > \pi(A)$ by

$$\Psi(x, K \setminus A) = \Psi(1 - x, A).$$

For any x, y with $0 < x, y$ and $x + y \le 1$, we define

$$\Psi(x, y) = \inf_{A,C} \{Q(A, C) : A \cap C = \emptyset; \pi(A) = x; \pi(C) = y\}.$$

It is not hard to see that the "inf" in these definitions is attained, so we could write "min" instead.

We also use another quantity related to conductance. Define the *spread* of a measurable subset A of K as

$$\psi(A) = \int_0^{\pi(A)} \Psi(x, K \setminus A)\, dx.$$

It is sometimes appropriate to consider the *global spread* or *modified spread*:

$$\psi_{gl}(A) = \int_0^1 \Psi(x, K \setminus A)\, dx\,, \qquad \psi_{mod}(A) = \pi(A) \int_0^1 \frac{\Psi(x, K \setminus A)}{\min\{x, 1-x\}}\, dx\ .$$

These are related by $\psi(A) \le \psi_{mod}(A) \le 4\psi_{gl}(A)$. The first inequality follows easily. For the second, $\Psi(x, K \setminus A)\, \pi(A)/x \le 2\Psi(x, K \setminus A)$ when $x \in [\pi(A)/2, 1/2]$. When $x \in (0, \pi(A))$ then $\Psi(x, K \setminus A)/x$ is non-decreasing and so $\int_0^{\pi(A)/2} \Psi(x, K \setminus A)/x\, dx \le \int_{\pi(A)/2}^{\pi(A)} \Psi(x, K \setminus A)/x\, dx$. A similar argument holds for the case of $x > 1/2$.

Each quantity will have its uses: $\psi(A)$ is the easiest to use and will prove to be the most flexible of the three in Section 3.3, $\psi_{gl}(A)$ is best when there is very high node expansion (e.g., K_n), while $\psi_{mod}(A)$ seems to behave similarly to a modified log-Sobolev inequality.

The quantity $\psi(A)/\pi(A)$ is the expected value of the ergodic flow into $K \setminus A$, originating in the worst $B \subset A$ of size $t = \pi(B)$ chosen uniformly in $[0, \pi(A)]$. Another way of saying this is to consider a measure preserving map $f : A \to [0, \pi(A)]$ such that $d\pi(f^{-1}(t))\, P_{f^{-1}(t)}(K \setminus A)$ is a monotone decreasing function of t. Then

$$\psi(A) = \int_A f(x) P_x(K \setminus A)\, d\pi(x). \tag{2.3}$$

Yet another way of defining $\psi(A)$ is to choose a random point y in $K \setminus A$ from the stationary distribution, and make one step from y to get a point y'. Then

$$\psi(A) = \mathsf{E}(f(y'))\, \pi(A^c)$$

(where we take $f(y') = 0$ for $y' \in K \setminus A$).

Perhaps the most instructive way of looking at this new quantity is the following. If we think of the Markov chain as a random walk on a graph, then $Q(A)$ measures the edge expansion of the set A. The quantity $\psi(A)$ can be viewed as a mixture of edge expansion and node expansion. For a fixed $t \in [0, \pi(A)]$, let C be a subset of A with $\pi(C) = t$ minimizing $Q(C, K \setminus A)$, and let $B = A \setminus C$. Since $\Psi(t, K \setminus A)$ is trivially monotone increasing in t, we get

$$\psi(A) \ge (\pi(A) - t)\Psi(t, K \setminus A) = \pi(B) Q(C, K \setminus A). \tag{2.4}$$

Part of the ergodic flow out of A goes from B, and for this part we measure the node-expansion by $\pi(B)$; the other part of the ergodic flow we measure directly, by $Q(C, K \setminus A)$.

How does one define node expansion in a Markov chain? A first attempt would be to consider, for a given A, the measure of all states in A that can be reached in one step. But this would be too sensitive to transitions out of A with tiny probability. A more reasonable choice is the following. Define $\beta(A)$ as the minimum stationary measure of a set $B \subseteq A$ such that $Q(A \setminus B, K \setminus A) = \frac{1}{2} Q(A)$ (i.e., B captures half of the flow out of

A). Then applying (2.4) with $t = \pi(A) - \beta(A)$, we get

$$\psi(A) \geq \frac{1}{2}\beta(A)Q(A).$$

To relate the spread ψ and the ergodic flow Ψ, note that $\beta(A) \geq \frac{1}{2}Q(A)$ (since a set of stationary measure p can absorb an ergodic flow of at most p), and so

$$\psi(A) \geq \frac{1}{4}Q(A)^2. \tag{2.5}$$

The ergodic flow out of a set measures the probability that we step out of the set, starting from an *average* element of the set. For small sets, it is sometimes useful to measure the probability that we step out of the set from *every* point of the set. To this end, we introduce the "local spread"

$$\xi(A) = \inf\{P_u(K \setminus A) : \ u \in A\}, \qquad \xi(t) = \inf\{\xi(A) : \ \pi(A) = t\} \quad (0 \leq t < 1),$$

and the quantity

$$\pi_1 = \inf\{t : \ \xi(t) \leq 1/10\}.$$

Note that when $a \geq x$ then

$$\Psi(x, 1-a) \geq \xi(a)x.$$

The quantity π_1 will replace the minimum stationary probability π_0 in the continuous setting.

3. General mixing results

Let $\mathcal{M}$ be a continuous state Markov chain with no atoms. Let σ^t be the distribution after the t-th step of the random walk has just been executed, starting from distribution σ. We also consider the distribution ρ^m obtained as follows: we choose an integer $Y \in \{0, 1, \ldots, m-1\}$ randomly and uniformly, and stop after Y steps.

Suppose we walk for m steps and stop. If σ^0 is absolutely continuous w.r.t. π, then it is easy to see by induction that so is σ^k for all $k \geq 0$. Thus $d\rho^m/d\pi$ exists. Let $g = m\, d\rho^m/d\pi = \sum_{i=0}^{m-1} d\sigma_i/d\pi$. Then for every measurable set $S \subseteq K$,

$$\int_S g\, d\pi = m\rho^m(S) = \sigma^0(S) + \cdots + \sigma^{m-1}(S)$$

is the expected number of times the set S is visited during the first m steps of the random walk. In particular,

$$\int_K g\, d\pi = m.$$

Suppose $f : K \to [0, 1]$, is a measure-preserving 1-1 map so that $g(f^{-1}(\cdot))$ is a monotone increasing function on $[0, 1]$. (It is easy to see by standard arguments that such an f exists; intuitively, we have just ordered K into $[0, 1]$ in increasing order of g.)

Set

$$q(x, y) = Q(f^{-1}([0, \min(x, y)]), f^{-1}([\max(x, y), 1])).$$

$q(x, y)$ measures the flow "straddling" $f^{-1}([\min(x, y), \max(x, y)])$; in other words it is the flow when we "block" this set.

3.1. Mixing time estimates in terms of a parameter function

We start with a theorem that bounds the distance of ρ^m from the stationary distribution in terms of a parameter function h. Let us call a function $h : [0, 1] \to \mathbb{R}_+$ a *mixweight function* if

$$\int_0^1 h(y)q(x, y)\,dy \geq 2x(1 - x)$$

holds for all $x \in (0, 1)$.

Theorem 3.1. *If h is a mixweight function, then for every starting distribution that is absolutely continuous with respect to π,*

$$d(\rho^m, \pi) \leq \frac{1}{m} \int_0^1 h(x)\,dx.$$

The ordering f is of course difficult to get a hold of in general. Instead, one can use :

Proposition 3.2. *If $h : [0, 1] \to \mathbb{R}_+$ satisfies $h(1 - y) = h(y)$ and*

$$\int_0^1 h(y)\Psi(y, K \setminus A)\,dy \geq 2\pi(A)(1 - \pi(A)) \tag{3.1}$$

for all A with $\pi(A) \leq 1/2$, then h is a mixweight function.

This follows easily by using the comparison $q(x, y) \geq \Psi(y, K \setminus f^{-1}([0, x]))$ when $x \leq 1/2$, while the case when $x > 1/2$ follows from $q(x, y) \geq \Psi(1 - y, K \setminus f^{-1}([x, 1]))$ and the symmetry $h(1 - y) = h(y)$.

It is not hard to see that there is always a mixweight function h with minimum integral (thus giving the best possible bound), but in general it will not be expressible by a formula. We'll return to the question of finding the "right" function h in section 3.3. Roughly speaking, we can think of $1/h(x)$ as a measure of expansion of sets of stationary measure approximately equal to x.

If we assume a stronger smoothness property of the starting distribution, we get a better bound on the mixing time, where the (often troublesome) ends of the integration can be omitted.

Theorem 3.3. *Suppose that the starting distribution σ satisfies $\sigma \leq M\pi$ for some $M > 0$. Let $0 < c < 1/2$ and let $h : [0, 1] \to \mathbb{R}_+$ satisfy*

$$\int_c^{1-c} h(y)q(x, y)\,dy \geq 2x(1 - x)$$

for every x with $c \leq x \leq 1 - c$. Then for every $m > 0$,

$$d(\rho^m, \pi) \leq 2cM + \frac{1}{m} \int_{1/M}^{1-1/M} h(x)\,dx.$$

The two results above are not quite satisfactory, since in many applications we are interested in random walks starting from a fixed state. However, in this case we meet a new difficulty. Even if the chain has a large conductance, it may happen that from the starting state s (which has measure 0) there is no escape (P_s is concentrated on s). Even if we assume that the Markov chain mixes from every starting point, we can adjust the exit distribution for the starting state so that it takes arbitrarily long to escape, without changing the conductance function. Thus in the case of continuous state Markov chains, we also need to take into account the "very local" expansion, using the quantity $\xi(0)$. We generalize slightly by considering $\xi(a)$, as this will prove useful later.

Theorem 3.4. *Suppose that $\xi(a) > 0$ and for some $c \le a$ let $h : [c, 1-c] \to \mathbb{R}_+$ satisfy $h(1-y) = h(y)$ and*

$$\int_c^{1-c} h(y)\Psi(y, K \setminus A)\, dy \ge 2\pi(A)(1-\pi(A))$$

for all A with $a \le \pi(A) \le 1/2$. Then for every $m > 1$ and every starting state,

$$d(\rho^m, \pi) \le \frac{8 \ln m}{m\xi(a)} + \frac{4}{m} \int_c^{1/2} h(x)\, dx.$$

3.2. Faster convergence to stationarity

General results [1, 17] imply that if starting at any state, $d(\sigma^m, \pi) < 1/4$ for some m, then $d(\sigma^{km}, \pi) < 1/2^k$. One might expect an exponential drop in the distance $d(\rho^m, \pi)$, but since ρ^m includes the term σ^0/m (see definition (1.1)), a rate of $1/m$ is the best that can be asserted. However, if we do a "better" averaging, then the exponential rate of drop does hold. For example, the general results mentioned above imply the following.

Corollary 3.5. *Under the conditions of Theorem 3.4, let*

$$R = 50 \frac{1}{\xi(0)} \ln \frac{1}{\xi(0)} + 32 \int_0^{1/2} h(x) dx,$$

and let Y be a binomially distributed random integer with expectation $1000kR$. From any starting point, run the chain for Y steps and stop. Then the distribution ρ of the last point satisfies

$$d(\rho, \pi) \le 2^{-k}.$$

This corollary implies that doing $O(R \ln(1/\delta))$ steps (where the exact number is a random variable easily generated), we get closer to the stationary distribution than δ.

3.3. Constructing mixweight functions

In this section we describe constructions that provide functions satisfying (3.1). Recall that for each of these, $1/h(x)$ is a measure of expansion of sets of size about x.

Considering (2.5) and Theorem 1.2, it would be natural to try something like

$$h(y) = \max_{\pi(A)=y} \frac{\pi(A)}{\psi(A)}.$$

For technical reasons, we have to smooth out this function and consider

$$\begin{aligned} h(y) &= \max_{y\le\pi(A)\le\min\{2y,1/2\}} \frac{4\pi(A)}{\psi(A)} \\ h_{gl}(y) &= \max_{0<\pi(A)\le\min\{2y,1/2\}} \frac{4\pi(A)}{\psi_{gl}(A)} \\ h_{mod}(y) &= \max_{0<\pi(A)\le\min\{2y,1/2\}} \frac{4\pi(A)^2}{y\,\psi_{mod}(A)} \end{aligned} \tag{3.2}$$

when $y \le 1/2$, and $h(y) = h(1-y)$ for $y > 1/2$. To show, for instance, that $h_{gl}(y)$ is a valid mixweight function we wish to apply Proposition 3.2. To this end, let A be a measurable set with $\pi(A) = x$ and $x \le 1/2$. Then

$$\begin{aligned} \int_0^1 \Psi(y, K\setminus A)\, h_{gl}(y)\, dy &\ge \int_{x/2}^{x} + \int_{x}^{1-x/2} \Psi(y, K\setminus A)\, h_{gl}(y)\, dy \\ &\ge \frac{4\pi(A)}{\psi_{gl}(A)} \left(\int_{x/2}^{x} + \int_{x}^{1-x/2} \Psi(y, K\setminus A)\, dy \right) \\ &\ge \frac{1}{2}\frac{4\pi(A)}{\psi_{gl}(A)} \left(\int_{0}^{x} + \int_{x}^{1} \Psi(y, K\setminus A)\, dy \right) \\ &= 2x \ge 2x(1-x). \end{aligned}$$

The same proof, but with only the $\int_\cdot^x$ terms implies that the $h(y)$ of (3.2) is also a mixweight function. A similar proof applies to $h_{mod}(y)$, but with the added observation that because $\Psi(y, K\setminus A)/y$ is increasing for $y \le \pi(A)$ then $\int_0^{x/2} \frac{\Psi(y,K\setminus A)}{y}\, dy \le \int_{x/2}^{x} \frac{\Psi(y,K\setminus A)}{y}\, dy$.

Each quantity is appropriate in different circumstances. Indeed, by an argument similar to (3.8) below, $h(y) \le \frac{16}{y} \max_{x\in[y,2y]} \frac{x^2}{\Psi(x)^2}$ while $h_{mod}(y) \le \frac{12}{y} \max_{x\le 2y} \frac{x^2}{\Psi(x)^2}$ and $h_{gl}(y) \le 8 \max_{x\le 2y} \frac{x}{\Psi(x)^2}$. The ability to restrict the domain of x makes $h(y)$ and $h_\ell(y)$ (see below) best when there are bottlenecks, i.e. tight constrictions that it is hard to pass through, as in the dumbbell graph discussed in the introduction. Experience shows that $\Psi(x)^2$ is typically of order $x^2\, polylog(1/x)$, which suggests that when node-expansion is moderate then $h_{mod}(y)$ will be the better of the global quantities, although $h_{gl}(y)$ can be best when there is very high node-expansion. For example, the quantity $h(y)$ in (3.2) implies mixing in time $O(\log n)$ for the lazy random walk on the complete graph K_n, while $h_{mod}(y)$ gives $O(\log\log n)$ and $h_{gl}(y)$ yields $O(1)$, the first isoperimetric proof of the correct mixing time bound.

Since the formulas in (3.2) are still pretty complicated, further specialization is in order. Using the lower bounds on $\psi(A)$ given in section 2, we get that each of the following

functions can serve as h:

$$h(y) = \max_{y \le \pi(A) \le \min\{2y, 1/2\}} \min_t \frac{4\,\pi(A)}{(\pi(A) - t)\Psi(t, K \setminus A)} \tag{3.3}$$

$$h(y) = \max_{y \le \pi(A) \le \min\{2y, 1/2\}} \frac{8\,\pi(A)}{\beta(A) Q(A)} \tag{3.4}$$

$$h(y) = \max_{y \le \pi(A) \le \min\{2y, 1/2\}} \frac{16\,\pi(A)}{Q(A)^2} \tag{3.5}$$

We can make one more refinement of these bounds which is sometimes useful. Let $\ell : [0, 1/2] \to \mathbb{R}$ be any function satisfying

$$0 < \ell(x) \le x \qquad \forall x,\ 0 \le x \le 1/2, \tag{3.6}$$

and the following semi-Lipschitz condition:

$$\ell(x) - \ell(y) \le x - y \qquad \forall x, y,\ 0 \le y \le x \le 1/2. \tag{3.7}$$

Then for $0 \le y \le 1/2$,

$$\begin{aligned} h_\ell(y) &= \sup_{\pi(A) \in \left(y, \min\left\{y + \ell(y), \frac{1}{2}\right\}\right)} \frac{4\,\pi(A)}{\psi_\ell(A)} \\ \textit{where } \psi_\ell(A) &= \int_{\pi(A) - \ell(\pi(A))}^{\pi(A)} \Psi(t, K \setminus A)\, dt \end{aligned}$$

and $h_\ell(1-y) = h_\ell(y)$, is a valid choice for h ($h(y)$ of (3.2) is just the case when $\ell(x) = x$). To see this, we proceed on the same lines as the proof for (3.2) :

$$\begin{aligned} \int_0^1 h_\ell(y)\, \Psi(y, K \setminus A)\, dy &\ge \int_{x - \frac{\ell(x)}{2}}^{x} h_\ell(y)\, \Psi(y, K \setminus A)\, dy \\ &\ge 2x \ge 2x(1 - x), \end{aligned}$$

using the fact that when $y \in [x - \frac{\ell(x)}{2}, x]$, we have $y \le x \le y + \ell(y)$ using the Lipschitz condition and so, $h_\ell(y) \ge 4\pi(A)/\psi_\ell(A)$.

The same argument holds for $\tilde{h}_\ell(y)$ and $\tilde{\psi}_\ell(A)$ given by $\tilde{\psi}_\ell(A) = \psi_\ell(A) + \psi_\ell(K \setminus A)$, with $h_\ell(y)$ the supremum over $\pi(A) \in [y - \ell(y), y + \ell(y)]$.

Corollary 3.6. *For every function ℓ satisfying (3.6) and (3.7), for every $m > 0$, and every starting distribution that is absolutely continuous with respect to π, we have*

$$d(\rho^m, \pi) \le \frac{2}{m} \int_0^{1/2} \tilde{h}_\ell(x)\, dx.$$

The function ℓ is nicer to have as a parameter than the function h, because the conditions on it do not depend on the Markov chain; but one would like to see further specialization.

One special choice of ℓ is $\ell(x) = \Psi(x)/4$ (which is legal by (2.2) and lemma 2.1). We have

$$\psi_\ell(A) \ge \ell(x)\, \Psi(x - \ell(x), 1 - x) + \ell(x)\, \Psi(x, 1 - x - \ell(x)) \ge \frac{3}{8}\, \Psi(x)^2. \tag{3.8}$$

Indeed, let $A, C \in \mathcal{A}$, $\pi(A) = x$, $\pi(C) = x - \Psi(x)/4$. Then $Q(A, K \setminus A) \geq \Psi(x)$. But $Q(A \setminus C, K \setminus A) \leq \pi(A \setminus C) = \Psi(x)/4$, and hence $Q(C, K \setminus A) \geq 3\,\Psi(x)/4$. Moreover, $\Psi(y) \geq 3\Psi(x)/4$ and $y \leq 5x/4$ in the range of y. It follows that $\tilde{h}_\ell(x) \leq 24x/\Psi(x)^2$ is a mixweight function. Applying Theorem 3.4, we get a continuous version of the main result of [14]:

Corollary 3.7. *Suppose that $\pi_1 > 0$. Then for every $m > 1$ and every starting state we have*

$$d(\rho^m, \pi) \leq \frac{425 \ln m}{m} + \frac{100}{m} \int_{\pi_1}^{1/2} \frac{x}{\Psi(x)^2}\, dx.$$

We can get a slightly sharper, but less explicit bound by considering a function h that satisfies the following condition for every measurable set A with $\pi(A) \leq 1/2$:

$$\begin{cases} \text{either} \quad h(y) \geq \dfrac{4\pi(A)}{\int_0^{\pi(A)} \Psi(t, K \setminus A)\, dt} & \text{for all } y \in [\frac{1}{2}\,\pi(A), \pi(A)], \\ \text{or} \quad h(y) \geq \dfrac{4\pi(A)}{\int_{\pi(A)}^{2\pi(A)} \Psi(t, K \setminus A)\, dt} & \text{for all } y \in [\pi(A), \frac{3}{2}\,\pi(A)]. \end{cases}$$

3.4. Explicit conductance bounds

The integral in the estimate in Theorem 3.1 may be difficult to evaluate (even if we use the simpler expressions in its corollaries). The following corollaries of Theorem 3.4 may be easier to apply. Each of these is derived by straightforward integration. First, let us note that by knowing a lower bound on the conductance, we get the following bound on the mixing time (which is closely related to the theorem of Jerrum and Sinclair, see section 3.5):

Corollary 3.8. *Suppose that for some $a, A > 0$, $\xi(a) > 0$, and an inequality of the form*

$$\Psi(x) \geq Ax \tag{3.9}$$

holds for all $a \leq x \leq 1/2$. Then for every starting distribution,

$$d(\rho^m, \pi) \leq \frac{8 \ln m}{m\xi(a)} + \frac{32}{mA^2} \ln \frac{1}{a}.$$

If we have a stronger bound, which expresses that smaller sets have better conductance, then we can save the logarithmic factor:

Corollary 3.9. *Consider a Markov chain such that $\xi(a) > 0$ for some $a > 0$. Suppose that for some $B > 0$, an inequality of the form*

$$\Psi(x) \geq Bx \ln(1/x) \tag{3.10}$$

holds for all $a \le x \le 1/2$. *Then*

$$d(\rho^m, \pi) \le \frac{8 \ln m}{m\xi(a)} + \frac{100}{mB^2}.$$

We get a stronger inequality (which in fact will be needed later on) if we can combine (3.9) and (3.10):

Corollary 3.10. *Consider a Markov chain with* $\xi(a) > 0$. *Suppose that for some* $A, B > 0$, *an inequality of the form*

$$\Psi(x) \ge \min\{Ax, Bx \ln(1/x)\}$$

holds for all $a \le x \le 1/2$. *Then*

$$d(\rho^m, \pi) \le \frac{8 \ln m}{m\xi(a)} + \frac{1}{m}\left(\frac{32}{A^2} \ln \frac{1}{a} + \frac{100}{B^2}\right).$$

Finally, we formulate another corollary, which can be considered as a continuous version of an inequality proved by Ledoux [13], using log-Sobolev inequalities:

Corollary 3.11. *Consider a Markov chain such that for some* $a, B > 0$ *we have* $\xi(a) > 0$, *and*

$$\Psi(x) \ge Bx\sqrt{\ln(1/x)}$$

for every $a < x < 1/2$. *Then*

$$d(\rho^m, \pi) \le \frac{8 \ln m}{m\xi(a)} + \frac{34}{mB^2} \ln \ln \frac{2}{a}$$

3.5. Finite chains

In the studies above, we made use of the atom-free nature of the Markov chain. If we want to extend our results to discrete Markov chains, we could reproduce the arguments, with some loss for "rounding" (since not all sizes occur as sizes of sets). A related, but somewhat weaker, argument is carried out in [18].

Here we sketch another possibility for obtaining analogous results for finite chains: One can replace each atom by an atom-free set of equivalent states. Consider a finite Markov chain $\mathcal{M}$ with state space $V = \{1, \dots, n\}$, transition probabilities p_{ij}, and stationary distribution π. Let S_i to be the interval $[\pi_1 + \cdots + \pi_{i-1}, \pi_1 + \cdots + \pi_i)$. We define a Markov chain $\hat{\mathcal{M}}$ on the unit interval $[0, 1)$ as follows: if we are at a point $x \in S_i$, then we choose an interval S_j with probability p_{ij}, choose a random point $y \in S_j$ uniformly, and move to y. Clearly this way we get a Markov chain whose stationary distribution is the Lebesgue measure on $[0, 1)$.

It is clear that no matter where we start, after the first step of $\hat{\mathcal{M}}$, the distribution of the current state will be uniform on each S_i, and the probability of being in S_i is the same as the probability of being at i in $\mathcal{M}$. Hence the mixing time of $\mathcal{M}'$ is the same as the mixing time of $\mathcal{M}$, and we can apply the previous bounds.

The difficulty with this method is that various expansion measures for the new chain

$\hat{\mathcal{M}}$ are only implicitly defined, and even though one can express them in terms of the original chain $\mathcal{M}$, these expressions will look quite artificial. As an example, consider the conductance function $\hat{\Psi}(x)$ of the new chain. It is easy to see that we have

$$\hat{\Psi}(x) = \min\{\sum_{i,j} Q_{ij} t_i (1 - t_j) : \ 0 \le t_i \le 1, \sum_i \pi_i t_i = x\},$$

but the minimum of this quadratic function is not easy to compute. (This is not surprising from the complexity point of view, since the standard conductance itself is NP-hard.) However, one can usually show that this conductance function is close to the conductance function of the finite chain. Since the details are cumbersome, we leave the detailed discussion of the finite case for another paper.

3.6. Further examples

As an illustration, we analyze the random walk on the dumbbell graph; this consists of two complete graphs of the same size $n/2$, connected by an edge. It is not hard to see that the mixing time is $\Theta(n^2)$. The conductance is $\Theta(1/n^2)$, so the Jerrum-Sinclair bound of $\Theta(n^4 \log n)$, is way off.

The reason why we can gain by applying Corollary 3.7 is that the only cut with really small conductance is the one separating the two complete graphs. Let $x < 1/2$, then it is not hard to see that

$$\hat{\Psi}(x) \ge \max\left\{\frac{1}{n^2},\, x\left(\frac{1}{2} - x\right)\right\},$$

whence integration gives the bound of $\Theta(n^2)$. This shows that we not only get rid of the start penalty, but also save the squaring of the conductance!

Benjamini and Mossel [3] considered a more interesting example, a Markov chain on the giant component of a super-critical percolation on the grid $\{-n, \ldots, n\}^d$. This is a random walk on the graph in which each edge in $\{-n, \ldots, n\}^d$ is retained with probability p. As $n \to \infty$ it is known that if p is greater than some threshold $p_c(\mathbb{Z}^d)$, then this graph will have a single "giant component" and many small components. They found that for every set $A \subset V(G)$ with $x = \pi(A) \le 1/2$ that the walk on the giant component has $\Psi(A) \ge c_{p,d}\, x^{1-1/d}/n$, where $c_{p,d}$ is a constant which depends on p and d. This is large enough to deal with the start penalty, and therefore, ignoring factors of d, a discretized version of our results would imply mixing in time $\tau = \Theta(n^2)$, the correct bound. Previous isoperimetric proofs could only show $\tau = O(n^2 \log n)$.

A problem in which the node-expansion (via spread ψ) proves useful is the lazy random walk on the cube $\{0,1\}^d$. In this walk, at each step a coordinate $i \in \{1 \ldots d\}$ is chosen uniformly at random and the value in this coordinate is set to 0 or 1 with probability $1/2$ each. It is well known that $\tau = \Theta(d \log d)$, but isoperimetric proofs are typically far from this. For instance, the conductance $\Phi = 1/d$ (take a half-cube by fixing coordinate d at $x_d = 0$), whereas Corollary 4.4 can be used to show that $\Psi(x) = \Omega(x \log(1 + 1/x))$. By Corollary 3.7 these imply mixing times of $\tau = O(d^3)$ or $\tau = O(d^2)$ respectively. Once again, high expansion on small sets suffices to deal with the start penalty.

However, observe that every subset $A \subset \{0,1\}^d$ will have a large fraction of its vertices

on the boundary. In fact when $\pi(A)$ is small then almost all vertices in A are boundary vertices (consider Hamming balls).

Given a set $A \subset \{0,1\}^d$ we lower bound $\psi(A)$ by using $\psi(A) = \int_{[0,\pi(A)]} t\, P_{f^{-1}(t)}(K \setminus A)\, dt$, as in (2.3). Then

$$\begin{aligned}
\mathsf{Q}_2(A) &:= \int_A \sqrt{\mathsf{P}_y(K \setminus A)}\, d\pi(y) \\
&= \int_0^x \sqrt{P_{f^{-1}(t)}(K \setminus A)}\, dt \\
&\le \mathsf{Q}_2(A)/2 + \sqrt{\int_{\mathsf{Q}_2(A)/2}^x t\, P_{f^{-1}(t)}(K \setminus A)\, dt \int_{\mathsf{Q}_2(A)/2}^x \frac{dt}{t}} \\
&\le \mathsf{Q}_2(A)/2 + \sqrt{\psi(A)\, \log(2x/\mathsf{Q}_2(A))}\ ,
\end{aligned}$$

where the first inequality follows from Cauchy-Schwartz. Solving for $\psi(A)$ gives

$$\psi(A) \ge \frac{1}{4}\, \frac{\mathsf{Q}_2(A)^2}{\log(2x/\mathsf{Q}_2(A))}\ .$$

Talagrand [22] showed that for this Markov chain the quantity

$$\mathsf{Q}_2(A) \ge \frac{1}{4}\, \pi(A)\, \sqrt{\frac{\log(2/\pi(A))}{d}}\ .$$

It follows from a discretized version of (3.2) that the Markov chain mixes in $\tau = O(d \log^2 d)$, very near the correct bound.

This method of "discrete gradients" is explored further in [19]. The quantity $\mathsf{Q}_2(A)$ is connected to concentration inequalities and the logarithmic Sobolev constant, the study of which was the goal of Talagrand's work.

4. Random walks in convex bodies

4.1. A logarithmic Cheeger inequality

A crucial ingredient of the proofs of rapid convergence for convex bodies is an isoperimetric inequality which asserts a lower bound on the surface area of a hypersurface dividing a convex body into two parts in terms of the product of the volumes of the two parts. Since it is technically more cumbersome to work with surface areas, a version of the isoperimetric inequality expands this hypersurface into a full-dimensional set. We use the notation that for any two sets S_1, S_2, $d(S_1, S_2)$, the distance between S_1 and S_2 is the infinum of distances between pairs of points one in S_1 and the other in S_2.

The following theorem was proved in [5] (improving results of [6, 15]):

Theorem 4.1. *Let $K \subset \mathbb{R}^n$ be a convex body of diameter D, and let $K = S_1 \cup S_2 \cup S_3$ be a partition into three measurable sets. Let $\varepsilon = d(S_1, S_2)$. Then*

$$\mathrm{vol}(S_3) \ge \frac{4\varepsilon}{D} \frac{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}{\mathrm{vol}K}.$$

Letting $\varepsilon \to 0$ yields the following:

Theorem 4.2. *Let $K \subset \mathbb{R}^n$ be a convex body of diameter D, and let K be separated into two parts S_1 and S_2 by a hypersurface F. Then*

$$\mathrm{vol}_{n-1}(F) \geq \frac{4}{D}\frac{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}{\mathrm{vol}K}.$$

A narrow cylinder of length D shows that the bound is sharp (at least in terms of these parameters): we take F the plane cut bisecting, and orthogonal to, the axis of the cylinder. Our main geometric tool is the following theorem, which gives a sharper bound in the case when S_1 or S_2 is small in volume relative to K.

Theorem 4.3. *Let $K \subset \mathbb{R}^n$ be a convex body of diameter D, and let $K = S_1 \cup S_2 \cup S_3$ be a partition into three measurable sets. Let $\varepsilon = d(S_1, S_2)$. Then*

$$\mathrm{vol}(S_3) \geq \frac{\varepsilon}{D}\frac{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}{\mathrm{vol}K}\ln\left(1 + \frac{\mathrm{vol}^2 K}{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}\right)$$

Again, applying this inequality to the ε-neighborhood of a surface cutting K into two parts, we get:

Corollary 4.4. *Let $K \subset \mathbb{R}^n$ be a convex body of diameter D, and let K be separated into two parts S_1 and S_2 by a hypersurface F. Then*

$$\mathrm{vol}_{n-1}(F) \geq \frac{1}{D}\frac{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}{\mathrm{vol}K}\ln\left(1 + \frac{\mathrm{vol}^2 K}{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}\right).$$

We remark that this version is sharp as $\mathrm{vol}(S_1)/\mathrm{vol}(K) \to 0$ (up to a factor of $1 + o(1)$ as $n \to \infty$). This is a consequence of a result of [20].

Fix constants c, $A > 0$ and let K be the n-dimensional truncated cone given as the solid of revolution of $f_n(t) = A\,(1 + c\,t/n)$ around the interval $t \in [0, 1]$. Then as $n \to \infty$ the cross sectional area of this solid becomes proportional to $(1 + c\,t/n)^{n-1} \xrightarrow{n\to\infty} e^{ct}$. If $s = c^{-1}\log c$ and S_1 is the region of the solid of revolution around the interval $[0, s]$, then

$$\frac{\mathrm{vol}_{n-1}(\partial S_1)}{\mathrm{vol}K} \xrightarrow{n\to\infty} \frac{c^2}{e^c - 1} \quad \text{and} \quad x = \frac{\mathrm{vol}_n(S_1)}{\mathrm{vol}K} \xrightarrow{n\to\infty} \frac{c-1}{e^c - 1}\,.$$

As $A \to 0^+$ then diameter $D \to 1$, and as $c \to \infty$ then $x \to 0^+$ and the ratio of $\mathrm{vol}_{n-1}(\partial S_1)/\mathrm{vol}_n(K)$ to the lower bound in Corollary 4.4 goes to 1.

4.2. Mixing of the ball walk

Suppose that K is a convex set in $\mathbf{R}^n$ with diameter at most D, containing the unit ball $B = B(0, 1)$. We consider the ball walk in K with step size δ: in this walk, at every step, we go from the current point $x \in K$ to a random point in $B(x, \delta) \cap K$ where $B(x, \delta)$ is the ball of radius δ with center x. This is called a "proper move". (In practice, to implement a proper move, one has to pick a random point from $B(x, \delta)$ and reject if it is not in K.) We assume that $\delta < D/100$. We call this walk the *proper move walk*. The distribution of

a step of the proper walk from $x \in K$ is given by

$$P_x(S) = \frac{\mathrm{vol}(S \cap B(x,\delta))}{\mathrm{vol}(K \cap B(x,\delta))}.$$

The steady state distribution of the proper move walk is not uniform, but is close enough to it that it helps us sample from the uniform (see [12] for details). To be more precise, we define the *local conductance* by

$$\ell(x) = \mathrm{vol}(K \cap B(x,\delta))/\mathrm{vol}(B(x,\delta))$$

for $x \in K$, and

$$E = \int_K \ell(x)\,dx.$$

Then the stationary distribution π of the proper move walk has density function ℓ/E.

To be able to apply Theorem 3.1 (or, in fact, Corollary 3.10) to this walk, we will have to estimate its local spread $\xi(a)$ (or π_1) and its conductance function Ψ.

Theorem 4.5. *The local spread of the proper move walk satisfies* $\pi_1 \geq \frac{1}{2}(\delta/D)^{2n}$.

Theorem 4.6. *For* $0 < x < 1/2$, *we have*

$$\Psi(x) > \min\left\{ \frac{1}{288\sqrt{n}}\,x,\ \frac{\delta}{81\sqrt{n}D}\,x\ln(1+1/x) \right\}.$$

4.3. Sampling from a convex body

Putting the results and methods of the previous two sections together, we obtain a better analysis of sampling algorithms in convex sets. Suppose K is a convex set in $\mathbf{R}^n$ with diameter at most D. Consider the ball walk in K as discussed in the previous section. In [12] it is shown that $O^*(n^2(\frac{D}{\delta})^2\ln(\frac{1}{\epsilon}))$ proper moves are sufficient to get within a total variation distance of ϵ of the steady state distribution starting from any state. The results of the last section plugged into this machinery immediately yield an improvement of this bound to $O((n(\frac{D}{\delta})^2 + n^2\ln(D/\delta))\ln(\frac{1}{\epsilon}))$, a savings of $O^*(n)$ when $\delta = O(1/\sqrt{n})$.

Further, [12] shows that after appropriate preprocessing, any convex body can be put into "isotropic" position where we have that the unit ball is contained in the body and most of the volume of the body is contained in a ball of radius $O(\sqrt{n})$. It was shown also in that paper that if we start already in the uniform distribution, then we may use $\delta = 1/\sqrt{n}$ and each proper move can be implemented in $O(1)$ amortized time as above. Thus in this case, it was known that if we start in the steady state in a body in isotropic position, after $O(n^3)$ steps, we get an almost uncorrelated sample. The results of this paper imply that we do not need the assumption about starting in the steady state; starting from any state, $O^*(n^3)$ steps suffice.

One remaining bottleneck in our algorithm is that it takes $O(n^5)$ steps to put a convex body in isotropic position. In a recent paper Lovász and Vempala [16] studied a volume algorithm based on a simulated annealing approach. Their Markov chain also requires $O(n^3)$ to obtain a sample, but only requires $O(n^4)$ steps to reach an approximation

to the isotropic position sufficient for their purposes. This leads to an aggregate $O(n^4)$ algorithm for approximating the volume.

5. Proofs I: Markov Chains

5.1. Proof of Theorem 3.1

Recall the ordering f and the function $q(\cdot,\cdot)$ defined in Section 3. We will identify A with $f(A)$ here, so that for example we will write $P_x((y,1])$ for $P_{f^{-1}(x)}(f^1((y,1]))$ etc. With this notation we have:

$$\frac{d}{dx}q(x,y) = \begin{cases} P_x((y,1]), & \text{if } x < y, \\ -P_x([0,y]), & \text{if } x > y. \end{cases} \tag{5.1}$$

For two measurable sets S, T (not necessarily disjoint), we denote by $F(S,T)$ the expected number of times we step from a state in S to a state in T during steps $1, 2, \ldots t$. Clearly

$$F(S,T) = \int_S g(u)P_u(T)d\pi(u).$$

Lemma 5.1. *For any $0 < t < 1$, we have*

$$\int_0^1 q(x,t)\,dg(x) = \sigma^m([0,t]) - \sigma^0([0,t]) = \sigma^0([t,1]) - \sigma^m([t,1]).$$

The inequality

$$|g(b) - g(a)| \le \frac{1}{q(b,a)}$$

follows easily, but we don't need it.

Proof. Clearly

$$F([0,t],(t,1]) = \int_0^t g(x)P_x((t,1])\,dx.$$

Using (5.1) for $x \le t$, we can integrate by parts and get

$$F([0,t],(t,1]) = \Big[g(x)q(x,t)\Big]_{x=0}^{t} - \int_0^t q(x,t)\,dg(x) = g(t)q(t,t) - \int_0^t q(x,t)\,dg(x).$$

Similarly, we get

$$F((t,1],[0,t]) = g(t)q(t,t) + \int_t^1 q(x,t)\,dg(x).$$

Using the equation

$$F((t,1],[0,t]) - F([0,t],(t,1]) = \sigma^m([0,t]) - \sigma^0([0,t])$$

(which can be derived by simple counting), the lemma follows. □

Let $t_0 = \inf\{t : g(t) \ge m\}$. We may assume that $g(t_0) = m$ (since g can be changed on a set of π-measure 0).

Lemma 5.2.

$$d(\rho^m, \pi) = \frac{1}{m} \int_0^{t_0} t \, dg(t) = \frac{1}{m} \int_{t_0}^1 (1-t) \, dg(t).$$

Proof. We have

$$d(\rho^m, \pi) = \rho^m([t_0, 1]) - \pi([t_0, 1]) = \int_{t_0}^1 \left(\frac{g(u)}{m} - 1 \right) du.$$

By partial integration we get

$$d(\rho^m, \pi) = \left[\left(\frac{g(t)}{m} - 1 \right) (t-1) \right]_{t_0}^1 - \frac{1}{m} \int_{t_0}^1 (t-1) \, dg(t).$$

Here the first term vanishes, whence we get the second expression in the lemma. The first follows similarly. □

Now we are ready to prove Theorem 3.1. By Lemma 5.1, we have for every $0 < y < 1$

$$\int_0^1 q(x, y) \, dg(x) = \sigma^0([y, 1]) - \sigma^m([y, 1]) \leq 1.$$

Multiplying both sides by $h(y)$ and integrating, we get

$$\int_0^1 h(y) \int_0^1 q(x, y) \, dg(x) \, dy \leq \int_0^1 h(y) \, dy.$$

On the left hand side we have

$$\int_0^1 h(y) \int_0^1 q(x, y) \, dg(x) \, dy = \int_0^1 \left(\int_0^1 h(y) q(x, y) \, dy \right) dg(x).$$

Here the expression in parenthesis is at least $2x(1-x)$ by the assumption on h. Thus we obtain that

$$\int_0^1 2x(1-x) \, dg(x) \leq \int_0^1 h(x) \, dx.$$

By Lemma 5.2

$$\begin{aligned} d(\rho^m, \pi) = \frac{1}{m} \int_0^{t_0} x \, dg(x) &\leq \frac{1}{2(1-t_0)m} \int_0^{t_0} 2x(1-x) \, dg(x) \\ &\leq \frac{1}{2(1-t_0)m} \int_0^1 h(x) \, dx. \end{aligned}$$

Similarly,

$$\begin{aligned} d(\rho^m, \pi) = \frac{1}{m} \int_{t_0}^1 (1-x) \, dg(x) &\leq \frac{1}{2t_0 m} \int_{t_0}^1 2x(1-x) \, dg(x) \\ &\leq \frac{1}{2t_0 m} \int_0^1 h(x) \, dx. \end{aligned}$$

Since either t_0 or $1 - t_0$ is at least $1/2$, the theorem follows.

5.2. Proof of Theorem 3.3

We have to modify the proof of Theorem 3.1. We define g, t_0 and q as before. The assumption on the density of σ^0 implies that the density of ρ^k is at most M, so we get that

$$g(t) \le mM \tag{5.2}$$

for almost all t. Since we can change g on a set of measure 0, we may assume that (5.2) holds for all t.

We can easily dispose of the case when $t_0 \ge 1 - c$:

$$d(\rho^m, \pi) = \rho^m([t_0, 1]) - \pi([t_0, 1]) \le M(1 - t_0) - (1 - t_0) \le (M - 1)c.$$

The case when $t_0 \le c$ is similarly easy. So suppose that $c < t_0 < 1 - c$.

Lemmas 5.1 and 5.2 remain valid, but we need the following variant of Lemma 5.2:

Lemma 5.3. *We have*

$$d(\rho^m, \pi) \le 2Mc + \frac{1}{m} \int_c^{t_0} t \, dg(t)$$

and

$$d(\rho^m, \pi) \le 2Mc + \frac{1}{m} \int_{t_0}^{1-c} (1 - t) \, dg(t).$$

Proof. Similarly as in the proof of lemma 5.2,

$$d(\rho^m, \pi) = \rho^m([t_0, 1]) - \pi([t_0, 1]) = \int_{t_0}^1 \left(\frac{g(u)}{m} - 1 \right) du.$$

This integral we split as

$$\int_{t_0}^1 \left(\frac{g(u)}{m} - 1 \right) du = \int_{t_0}^{1-c} + \int_{1-c}^1 . \tag{5.3}$$

For the first term we get by partial integration:

$$\int_{t_0}^{1-c} \left(\frac{g(u)}{m} - 1 \right) du = \left[\left(\frac{g(t)}{m} - 1 \right) (t - 1) \right]_{t_0}^{1-c} + \frac{1}{m} \int_{t_0}^{1-c} (1 - t) \, dg(t).$$

Here the first term can be estimated as follows:

$$\left(\frac{g(1 - c)}{m} - 1 \right) c \le (M - 1)c.$$

The second integral in (5.3) can be estimated easily:

$$\int_{1-c}^1 \left(\frac{g(u)}{m} - 1 \right) du \le (M - 1)c.$$

This proves the second part of the lemma; the first is similar. □

By Lemma 5.1, we have for every $0 < y < 1$

$$\int_0^1 q(x, y) \, dg(x) = \sigma^0([y, 1]) - \sigma^m([y, 1]) \le 1.$$

Multiplying both sides by $h(y)$ and integrating, we get

$$\int_c^{1-c} h(y) \int_c^{1-c} q(x,y)\, dg(x)\, dy \le \int_c^{1-c} h(y)\, dy.$$

On the left hand side we have

$$\int_c^{1-c} h(y) \int_c^{1-c} q(x,y)\, dg(x)\, dy = \int_c^{1-c} \left(\int_c^{1-c} h(y) q(x,y)\, dy \right) dg(x).$$

Here the expression in parenthesis is at least $2x(1-x)$ by the assumption on h. Thus we obtain that

$$\int_c^{1-c} 2x(1-x)\, dg(x) \le \int_c^{1-c} h(x)\, dx.$$

By Lemma 5.3

$$\begin{aligned} d(\rho^m, \pi) &\le 2Mc + \frac{1}{m} \int_c^{t_0} x\, dg(x) \le 2Mc + \frac{1}{2(1-t_0)m} \int_c^{t_0} 2x(1-x)\, dg(x) \\ &\le 2Mc + \frac{1}{2(1-t_0)m} \int_c^{1-c} h(x)\, dx. \end{aligned}$$

Similarly,

$$\begin{aligned} d(\rho^m, \pi) &\le 2Mc + \frac{1}{m} \int_{t_0}^{1-c} (1-x)\, dg(x) \le 2Mc + \frac{1}{2t_0 m} \int_{t_0}^{1-c} 2x(1-x)\, dg(x) \\ &\le 2Mc + \frac{1}{2t_0 m} \int_c^{1-c} h(x)\, dx. \end{aligned}$$

Since either t_0 or $1-t_0$ is at least $1/2$, the theorem follows.

5.3. Proof of Theorem 3.4

Let

$$n = \left\lceil \frac{\ln m}{\xi(0)} \right\rceil.$$

Consider the distribution σ^n. This is not absolutely continuous with respect to π in general, but we may decompose it as $\sigma^n = \sigma' + \sigma''$, where σ' is absolutely continuous with respect to π and σ'' is concentrated on a set K_0 with stationary measure 0. We may assume that

$$\mathsf{P}_x(K_0) = 0 \tag{5.4}$$

for every $x \notin K_0$. Indeed, define the sets $K_1, K_2, \ldots$ recursively by $K_i = K_{i-1} \cup \{x \in K : P_x(K_{i-1}) > 0\}$. Then it follows by induction that $\pi(K_i) = 0$: if this is true for i, then it follows for $i+1$ by the equation

$$0 = \pi(K_i) = \int_K P_x(K_i)\, d\pi(x).$$

Hence $\bar{K}_0 = \cup_i K_i$ has stationary measure 0, and we have $P_x(\bar{K}_0) = 0$ for $x \notin \bar{K}_0$. So we can replace K_0 by $\bar{K}_0$.

We claim that

$$\sigma^n(K_0) < \frac{1}{m}.$$

Indeed, for $k > 0$, we have

$$\begin{aligned}\sigma^k(K_0) &= \int_K P_x(K_0)\, d\sigma^{k-1}(x) \\ &= \int_{K_0} P_x(K_0)\, d\sigma^{k-1}(x) + \int_{K\setminus K_0} P_x(K_0)\, d\sigma^{k-1}(x).\end{aligned}$$

Here the second integral is 0 by (5.4), while the first one can be estimated, by the definition of $\xi(0)$, by

$$\int_{K_0} P_x(K_0)\, d\sigma^{k-1}(x) \le (1-\xi(0))\sigma^{k-1}(K_0).$$

Hence it follows by induction that $\sigma^k(K_0) \le (1-\xi(0))^k$. In particular,

$$\sigma^n(K_0) \le (1-\xi(0))^n \le e^{-n\xi(0)} \le \frac{1}{m}.$$

Consider a walk $(v^0, v^1, \dots, v^{m-1})$, and choose an integer $Y \in \{0, \dots, m-1\}$ uniformly. Call this trial *bad*, if either $v^n \in K_0$ or $Y < n$. The probability that this happens is at most $(1/m) + (n/m)$. If the trial is not bad, then its result v^Y can be considered as the result of starting from the distribution $\sigma'/\sigma'(K)$ (which is absolutely continuous with respect to π) and going for $Y - n$ steps (which is an integer uniformly distributed in $\{0, \dots, m-n-1\}$). By Theorem 3.1, this proves that

$$d(\rho^m, \pi) \le \frac{n+1}{m} + \frac{2}{m-n}\int_0^{1/2} h(x)\, dx\ .$$

When $x \le a$ then $h(\cdot)$ given in the Theorem can be extended to $[0,1]$ by setting $\tilde h(x) = 4/a\,\xi(a) + h(x)\,\delta(x \ge c)$. It is easily verified that whenever $\pi(A) \le a$ then $\int_0^a \tilde h(y)\,\Psi(y, K\setminus A)\, dy \ge 2\pi(A)(1-\pi(A))$. Then $\int_0^a \tilde h(x)\, dx = 4/\xi(a) + \int_c^a h(x)\, dx$.

6. Proofs II: geometry

6.1. Lemmas

We need the following "elementary" inequality.

Lemma 6.1. *Let $a < b < c < d$. Then*

$$\frac{d-a}{c-b} \ge \frac{(e^b - e^a)(e^d - e^c)}{(e^c - e^b)(e^d - e^a)} \ln\left(1 + \frac{(e^d - e^a)^2}{(e^b - e^a)(e^d - e^c)}\right)$$

Proof. The inequality is invariant under shifting all variables, so we may assume that $b = 0$. Then we can write the inequality as

$$\frac{(d-a)}{(e^d - e^a)}\frac{e^c - 1}{c} \ge u \ln\left(1 + \frac{1}{u}\right),$$

where

$$u = \frac{(1-e^a)(e^d-e^c)}{(e^d-e^a)^2}.$$

Now the left hand side is a monotone increasing function of c. The right hand side is monotone increasing as a function of u, but u is a monotone decreasing function of c. Thus the right hand side is a monotone decreasing function of c. It follows that it suffices to prove the inequality in the limiting case when $c \to 0$, in which case we get

$$d - a \geq \frac{(1-e^a)(e^d-1)}{e^d-e^a} \ln\left(1 + \frac{(e^d-e^a)^2}{(1-e^a)(e^d-1)}\right).$$

Substituting $x = e^d - 1$ and $y = 1 - e^a$, we have to prove that for $x, y > 0$,

$$\ln\frac{1+x}{1-y} \geq \frac{xy}{x+y} \ln\left(1 + \frac{(x+y)^2}{xy}\right),$$

Here $y \in (0,1)$ and so

$$\frac{1+x}{1-y} = 1 + \frac{x+y}{1-y} \geq 1 + x + y.$$

It suffices to prove that

$$\ln(1+x+y) \geq \frac{xy}{x+y} \ln\left(1 + \frac{(x+y)^2}{xy}\right),$$

or

$$(1+x+y)^{\frac{x+y}{xy}} \geq 1 + \frac{(x+y)^2}{xy},$$

which is clear as $(1+A)^B \geq 1 + B\,A$ when $A \geq 0$ and $B \geq 1$. □

This inequality generalizes to integrals of log-concave functions as follows.

Corollary 6.2. *Let $a < b < c < d$, and let F be a log-concave function on the interval $[a,d]$. Let $\nu(a,b) = \int_a^b F(t)\,dt$. Then*

$$\frac{d-a}{c-b} \geq \frac{\nu(a,b)\nu(c,d)}{\nu(b,c)\nu(a,d)} \ln\left(1 + \frac{\nu(a,d)^2}{\nu(a,b)\nu(c,d)}\right) \tag{6.1}$$

Proof. There is a (unique) exponential function $h(t) = \alpha e^{\gamma t}$ such that $F(b) = h(b)$ and $F(c) = h(c)$. Let $\mu(a,b) = \int_a^b h(t)\,dt$. Then

$$\mu(a,b) \geq \nu(a,b), \quad \mu(c,d) \geq \nu(c,d), \quad \text{but} \quad \mu(b,c) \leq \nu(b,c).$$

Hence elementary arithmetic gives that

$$\frac{\nu(a,b)\nu(c,d)}{\nu(a,d)} \leq \frac{\mu(a,b)\mu(c,d)}{\mu(a,d)}, \tag{6.2}$$

and

$$\frac{\nu(a,d)}{\nu(b,c)} \leq \frac{\mu(a,d)}{\mu(b,c)}. \tag{6.3}$$

If $\nu(a,d) \le \mu(a,d)$ then by applying (6.2) and that the function $x\ln(1+(1/x))$ is monotone increasing, we see that by replacing F by h we get a stronger inequality.

When $\nu(a,d) > \mu(a,d)$ then the same result holds, but this time write the right hand side of (6.1) as

$$\frac{\nu(a,d)}{\nu(b,c)}\frac{\nu(a,b)\nu(c,d)}{\nu(a,d)^2}\ln\left(1+\frac{\nu(a,d)^2}{\nu(a,b)\nu(c,d)}\right),$$

apply a modification of (6.2) with squares $\nu(a,d)^2$ and $\mu(a,d)^2$ in the denominators, and finish with (6.3).

But (6.1) for an exponential function is just lemma 6.1. □

We quote a similar lemma (Lemma 3.8) from [11]:

Corollary 6.3. *Let $a<b<c<d$, let F be a log-concave function on the interval $[a,d]$, and let g be a linear function on $[a,d]$. Let $\mu(a,b)=\int_a^b F(t)g(t)^{n-1}\,dt$. Then*

$$\mu(b,c) \ge \frac{1}{4\sqrt{n}}\frac{|F(b)-F(c)|}{\max\{F(b),F(c)\}}\min\{\mu(a,b),\mu(c,d)\}.$$

Finally, we need the following lemma, which easily follows from lemma 3.5 in [12]:

Lemma 6.4. *Let $x,y\in K$ with $\|x-y\|<\delta/\sqrt{n}$. Then*

$$d(P_x,P_y) < 1-\frac{\min\{\ell(x),\ell(y)\}}{4\max\{\ell(x),\ell(y)\}}.$$

6.2. Proof of theorem 4.5

Blowing up $K\cap B(x,\delta)$ by a factor of $(\delta/D)^n$ we get a set covering K, whence

$$\ell(x)=\frac{\mathrm{vol}(K\cap B(x,\delta))}{\mathrm{vol}(B(x,\delta))} \ge \left(\frac{\delta}{D}\right)^n \frac{\mathrm{vol}(K)}{\mathrm{vol}(B(x,\delta))}. \tag{6.4}$$

Let S be a set with stationary measure $\pi(S)\le\frac{1}{2}(\delta/D)^{2n}$, and let $x\in K$. Then (using (6.4) in the numerator and $\ell(x)\le 1$ in the denominator) we get

$$\begin{aligned}\pi(S) &= \int_S \ell(x)\,dx \Big/ \int_K \ell(x)\,dx \\ &\ge \left(\frac{\delta}{D}\right)^n \frac{\mathrm{vol}(S)}{\mathrm{vol}(B(x,\delta))}.\end{aligned}$$

Thus

$$\begin{aligned}P_x(S) &= \frac{\mathrm{vol}(S\cap B(x,\delta))}{\mathrm{vol}(K\cap B(x,\delta))} \le \frac{\mathrm{vol}(S)}{\ell(x)\mathrm{vol}(B(x,\delta))} \\ &\le \left(\frac{D}{\delta}\right)^{2n}\pi(S)\frac{\mathrm{vol}(B(x,\delta))}{\mathrm{vol}(K)} \le \frac{1}{2}.\end{aligned}$$

6.3. Proof of theorem 4.6

The proof follows the lines of the proof of Theorem 3.1 in [12]. Let

$$A = A(x) = \min\left\{\frac{x}{288\sqrt{n}}, \frac{\delta}{81\sqrt{n}D}x\ln(1+(1/x))\right\}$$

(this is the lower bound on $\Psi(x)$ we claim). It is easy to check that $A(x)$ is a monotone increasing concave function of x.

Let $K = S_1 \cup S_2$ be a partition of K into two non-empty parts, with $0 < x = \pi(S_1) \leq 1/2$. We want to show that

$$Q(S_1, S_2) > A.$$

Define, for $y \in K$,

$$h(y) = \begin{cases} P_y(S_1) & \text{if } y \in S_2, \\ P_y(S_2) & \text{if } y \in S_1, \end{cases}$$

(the probability that a random step from y crosses over). Since

$$\int_{S_1} h(y)\,d\pi(y) = Q(S_1, S_2) = Q(S_2, S_1) = \int_{S_2} h(y)\,d\pi(y),$$

it suffices to show that

$$\int_K h(y)\,d\pi(y) \geq 2A.$$

Suppose this is false. Let χ_1 denote the incidence function of S_1. Then, changing the integration measure to the Lebesgue measure, we have

$$\int_K h(y)\ell(y)\,dy < 2A\int_K \ell(y)\,dy.$$

The definition of x gives

$$\int_K \chi_1(y)\ell(y)\,dy = x\int_K \ell(y)\,dy$$

By the Localization Lemma (2.5 in [11]), there exist points a, b and a linear function $g(t) \geq 0$ for $t \in [0,1]$ such that setting $y(t) = (1-t)a + tb$, we have

$$\int_0^1 g(t)^{n-1}h(y(t))\ell(y(t))\,dt < 2A\int_0^1 g(t)^{n-1}\ell(y(t))\,dt \tag{6.5}$$

and

$$\int_{J_1} g(t)^{n-1}\ell(y(t))\,dt = x\int_0^1 g(t)^{n-1}\ell(y(t))\,dt, \tag{6.6}$$

where $J_i = \{t \in [0,1] : \ y(t) \in S_i\}$ for $i = 1, 2$. We abuse the notation by writing $h(t)$ and $\ell(t)$ for $h(y(t))$ and $\ell(y(t))$.

For $i = 1, 2$, define

$$J_i' = \{t \in J_i : h(t) < 1/16\},$$

and let

$$B = [0,1] \setminus (J_1' \cup J_2').$$

It is easy to see that J_i' and B are all measurable sets.

Define a measure μ on $I = [0,1]$ by

$$\mu(T) = \int_T g(t)^{n-1}\ell(t)\,dt$$

for any measurable subset T. We have $\mu(J_1) = x\mu(I)$ by (6.6). Furthermore,

$$\int_0^1 g(t)^{n-1}h(y(t))\ell(y(t))\,dt \geq \frac{1}{16}\int_B g(t)^{n-1}\ell(t)\,dt = \frac{1}{16}\mu(B),$$

and hence (6.5) implies that

$$\mu(B) < 32A\mu(I). \tag{6.7}$$

I. First we consider the crucial case when $J_1' = [0,r)$, $B = [r,s]$ and $J_2' = (s,1]$ are intervals. Then Corollary 6.2, applied to the log-concave function $g^{n-1}\ell$, implies that

$$\begin{aligned}
\frac{1}{s-r} &\geq \frac{\mu(J_1')\mu(J_2')}{\mu(B)\mu(I)}\ln\left(1+\frac{\mu(I)^2}{\mu(J_1')\mu(J_2')}\right)\\
&\geq \frac{(\mu(J_1)-\mu(B))(\mu(J_2)-\mu(B))}{\mu(B)\mu(I)}\ln\left(1+\frac{\mu(I)^2}{\mu(J_1)\mu(J_2)}\right)\\
&> \frac{(x-32A)(1-x-32A)}{32A}\ln\left(1+\frac{1}{x(1-x)}\right)\\
&> \frac{x}{81A}\ln\left(1+\frac{1}{x}\right) \geq \frac{D\sqrt{n}}{\delta}.
\end{aligned}$$

So it follows that $s-r < (\delta/\sqrt{n}D)$, and hence

$$|y(s)-y(r)| < \frac{\delta}{D\sqrt{n}}|b-a| \leq \frac{\delta}{\sqrt{n}}.$$

Choose $u < r$ and $v > s$ such that $v-u < \delta/(D\sqrt{n})$. Then $h(u), h(v) < 1/16$. On the other hand,

$$h(u)+h(v) = P_u(S_2)+P_v(S_1) = 1-P_u(S_1)+P_v(S_1) \geq 1-d(P_u,P_v),$$

and hence $d(P_u,P_v) \geq 7/8$. By Lemma 6.4, this implies that

$$\max\{\ell(u),\ell(v)\} > 2\min\{\ell(u),\ell(v)\}.$$

Letting u tend to r and v to s, then $\max\{\ell(r),\ell(s)\} \geq 2\min\{\ell(r),\ell(s)\}$ by the continuity of the function ℓ.

Now we can invoke Corollary 6.3, and get that

$$\mu(B) \geq \frac{1}{8\sqrt{n}}\min\{\mu(J_1'),\mu(J_2')\}.$$

Since

$$\mu(J_i') \geq \mu(J_i)-\mu(B) \geq x\mu(I)-\mu(B),$$

this implies that

$$\mu(B) \geq \frac{1}{9\sqrt{n}}x\mu(I),$$

which contradicts (6.7).

II. The general case can be reduced to the case of three intervals similarly as in [12]. Let $[x_i, y_i]$ be all maximal intervals contained in B. By the special case settled above, we have

$$\mu([x_i, y_i]) \geq A(\mu(T_i)),$$

where T_i is either $[a, x_i]$ or $[y_i, b]$. Summing over all i, we get

$$\mu(B) \geq \sum_i A(\mu(T_i)) \geq A(\mu(\cup_i T_i))$$

(since A is sub-additive). To conclude, it suffices to notice that either J_1' or J_2' is contained in $\cup_i T_i$. Indeed, for any $u \in J_1'$ and $v \in J_2'$ there is an interval $[x_i, y_i]$ separating them, and then either u or v is contained in T_i. □

6.4. Proof of Theorem 4.3

Suppose, by way of contradiction, that

$$\mathrm{vol}(S_3) < \frac{\varepsilon}{D} \frac{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}{\mathrm{vol}K} \ln\left(1 + \frac{\mathrm{vol}^2 K}{\mathrm{vol}(S_1)\mathrm{vol}(S_2)}\right)$$

Let $\nu > 0$ be a sufficiently small real number (which we will let go to zero later). Consider the functions

$$f(x) = \begin{cases} \frac{1}{\mathrm{vol}(S_1)-\nu} - \frac{1}{\mathrm{vol}(K)}, & \text{if } x \in S_1, \\ -\frac{1}{\mathrm{vol}(K)}, & \text{otherwise,} \end{cases}$$

and

$$g(x) = \begin{cases} \frac{1}{\mathrm{vol}(S_2)-\nu} - \frac{1}{\mathrm{vol}(K)} & \text{if } x \in S_2, \\ -\frac{1}{\mathrm{vol}(K)}, & \text{if } x \in S_1 \cup S_3. \end{cases}$$

Then clearly

$$\int_{\mathbb{R}^n} f > 0, \quad \int_{\mathbb{R}^n} g > 0.$$

By the Localization Lemma, there exist two points $u, v \in \mathbb{R}^n$ and a real number γ such that

$$\int_0^1 f(tu + (1-t)v)e^{\gamma t}\, dt \;>\; 0, \qquad \text{and} \tag{6.8}$$

$$\int_0^1 g(tu + (1-t)v)e^{\gamma t}\, dt \;>\; 0. \tag{6.9}$$

We may assume that $u, v \in K$ (else, we replace u and v by the endpoints of the intersection of the segment $[u, v]$ with K, which is trivially non-empty). Let

$$J_i = \{t \in [0,1] :\ ta + (1-t)b \in S_i\},$$

and

$$\mu(J) = \int_J e^{\gamma t}\, dt \Big/ \int_0^1 e^{\gamma t}\, dt.$$

Then (6.8) can be written as

$$\mu(J_1) > \frac{\mathrm{vol}(S_1) - \nu}{\mathrm{vol}(K)},$$

and (6.9), as

$$\mu(J_2) > \frac{\mathrm{vol}(S_2) - \nu}{\mathrm{vol}(K)}.$$

It follows that

$$\mu(J_3) = 1 - \mu(J_1) - \mu(J_2) < \frac{\mathrm{vol}(S_3) + 2\nu}{\mathrm{vol}(K)}.$$

We claim that

$$\mu(J_3) < \frac{\varepsilon}{D}\mu(J_1)\mu(J_2)\ln\left(1 + \frac{1}{\mu(J_1)\mu(J_2)}\right). \tag{6.10}$$

Indeed, by the monotonicity of the function $x\ln(1+1/x)$, we have

$$\frac{\varepsilon}{D}\mu(J_1)\mu(J_2)\ln\left(1 + \frac{1}{\mu(J_1)\mu(J_2)}\right) - \mu(J_3)$$

$$\geq \frac{\varepsilon}{D}\frac{(\mathrm{vol}(S_1) - \nu)(\mathrm{vol}(S_2) - \nu)}{\mathrm{vol}(K)^2}\ln\left(1 + \frac{\mathrm{vol}(K)^2}{(\mathrm{vol}(S_1) - \nu)(\mathrm{vol}(S_2) - \nu)}\right) - \frac{\mathrm{vol}(S_3) + 2\nu}{\mathrm{vol}(K)}.$$

Choosing ν sufficiently small, this will be positive by the indirect hypothesis. This proves (6.10).

In the intuitively hardest case when $J_1 = [0, b]$, $J_2 = [c, 1]$ and $J_3 = [b, c]$ for some $0 < b < c < 1$, this inequality directly contradicts lemma 6.1. The general case can be reduced to this case just like in [12]. □

References

[1] Aldous, D.J. and Diaconis, P. (1986) Shuffling cards and stopping times. *American Mathematical Monthly* **93** 333–348.

[2] Aldous, D.J., Lovász, L. and Winkler, P. (1997) Mixing times for uniformly ergodic Markov chains. *Stochastic Processes and their Applications* **71** 165–185.

[3] Benjamini, I. and Mossel, E. (2003) On the mixing time of simple random walk on the super critical percolation cluster. *Probability Theory and Related Fields* **125**:(3) 408–420.

[4] Diaconis, P. and Saloff-Coste, L. (1996) Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability* **6** 695–750.

[5] Dyer, M. and Frieze, A. (1992) Computing the volume of convex bodies: a case where randomness provably helps. In Béla Bollobás (editor), *Probabilistic Combinatorics and Its Applications*.

[6] Dyer, M., Frieze, A. and Kannan, R. (1989) A random polynomial time algorithm for estimating volumes of convex bodies. *Proceedings of the 21st Annual ACM Symposium on the Theory of Computing (STOC)* 375–381.

[7] Frieze, A. and Kannan, R. (1999) Log-Sobolev inequalities and sampling from log-concave distributions. *Annals of Applied Probability* **9**:(1) 14–26.

[8] Jerrum, M. and Sinclair, A. (1988) Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved. *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC)* 235–243.

[9] Jerrum, M. and Son, J. (2002) Spectral gap and log-Sobolev constant for balanced matroids, *Proceedings of the 31st Annual IEEE Symposium on Foundations of Computer Science (FOCS),* IEEE Computer Soc. 721–729.

[10] Jerrum, M., Son, J., Tetali, P. and Vigoda, E. (2004) Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Annals of Applied Probability* **14**:(4) 1741–1765.

[11] Kannan, R., Lovász, L. and Simonovits, M. (1995) Isoperimetric problems for convex bodies and a localization lemma. *Discrete and Computational Geometry* **13** 541–559.

[12] Kannan, R., Lovász, L. and Simonovits, M. (1997) Random walks and an $O^*(n^5)$ volume algorithm. *Random Structures and Algorithms* **11**:(1) 1–50.

[13] Ledoux, M. (1999) Concentration of measure and logarithmic Sobolev inequalities. *Séminaire de Probabilités* XXXIII, Lecture Notes in Math. **1709**, Springer, Berlin 120–216.

[14] Lovász, L. and Kannan, R. (1999) A logarithmic Cheeger inequality and mixing in random walks. *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC)*, 282-287.

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

[16] Lovász, L. and Vempala, S. (2003) Simulated Annealing in Convex Bodies and an $O^*(n^4)$ Volume Algorithm. *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)* 650–659.

[17] Lovász, L. and Winkler, P. (1995) Efficient stopping rules for Markov chains. *Proceedings of the 27th ACM Symposium on the Theory of Computing (STOC)* 76–82.

[18] Montenegro, R. (2002) Faster mixing by isoperimetric inequalities, Ph.D. Thesis, Department of Mathematics, Yale University.

[19] Montenegro, R. (2005) Vertex and edge expansion properties for rapid mixing. *Random Structures and Algorithms* **26**:(1–2) 52–68.

[20] Montenegro, R. (2005) A sharp isoperimetric inequality for convex bodies. *Israel Journal of Mathematics*, to appear.

[21] Morris, B. and Peres, Y. (2003) Evolving sets and mixing. *Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC)*, 279–286.

[22] Talagrand, M. (1993) Isoperimetry, Logarithmic Sobolev Inequalities on the Discrete Cube, and Margulis' Graph Connectivity Theorem. *Geometric and Functional Analysis.* **3**:(3) 295–314.