Abstract

Faster Mixing by Isoperimetric Inequalities

Ravi R. Montenegro

2002

This thesis is concerned with isoperimetric methods for studying the rate at which Markov chains approach their steady state distribution. We begin by proving a new isoperimetric bound on the mixing time using a quantity which we call blocking conductance $\phi(x)$, this is an extension of conductance Φ and average conductance $\Phi(x)$. We then look at the three methods for bounding conductance of which we are aware : geometry, induction, and canonical paths. We extend all three of these methods and obtain bounds on the blocking conductance $\phi(x)$ or the conductance function $\Phi(x)$; in all three cases these give significant improvements over conductance based bounds for the mixing time. We end by considering a new isoperimetric quantity $h_2^+(x)$; we prove a mixing time bound in terms of $h_2^+(x)$ and conjecture a stronger theorem which may give optimal mixing time bounds for a wide range of Markov chains including geometric, inductive, and product Markov chains.

Faster Mixing by Isoperimetric Inequalities

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Ravi R. Montenegro

Dissertation Director: professor Ravindran Kannan

December 2002

Contents

List of Figures

Acknowledgements

First of all I would like to thank the readers – Ravi Kannan, Lazlo Lovász and David Pollard – for putting in the time to read this dissertation.

I would not have made it to this place without the help of many people.

My Mathematical career has been most influenced by my advisor, Ravi Kannan. When I was uncertain of what to study he introduced me to the field of rapidly mixing Markov chains, and whenever work slowed he always had new ideas.

The few months I spent in Edinburgh with Mark Jerrum did a great deal to stimulate my interests. It was most enjoyable to have three people around with the same interests – Mark Jerrum, Eric Vigoda and Jung-Bae Son – and for a chance at the periodic talk about Markov chains.

There are many people who encouraged me to pursue the path of Mathematical studies. George Gilbert and Rhonda Hatcher organized a wonderful REU program at Texas Christian University, which stimulated me enough that I immediately decided that I just had to go to graduate school. At CalTech, Peter Ozsvath was unknowingly a mentor, both through his Algebraic Topology class and through office discussions about grad school. My first "research paper" with Professor Mario Martelli of Cal State Fullerton served to encourage me into studying the higher maths. Back at Troy High School, I'll never forget Mr. Knox's sense of humor while teaching Calculus. And in my younger years Kathy Griffen and Joan Carlson of Mendocino Middle School did a great job of encouraging all of our young minds to think about math in a fun way.

Of course, none of this would have been possible without my family. From a young age my parents Ricardo and Patricia Montenegro always encouraged me to do whatever I wanted (except when I ate my dad's plants). Those workbooks sure didn't go to waste! My studies took a great leap when my grandparents José and Zita Montenegro took care of me for a summer, driving me to CTY every day. It was always nice to know that my grandparents in Minnesota were admiring my accomplishments from afar. And Nessa, Alita and Vanea (unwillingly) put up with my tiring logical arguments for many a year.

These last few years, my wife Yuka Amano has been a great support. She crossed the sea to be with me, and has always been certain that I can do what I set my mind to. Stick to your studies, before you know it you'll be the one writing the acknowledgements. And no, I didn't forget Tama.

Summary

This thesis is concerned with isoperimetric methods for studying the rate at which Markov chains approach their steady state distribution. We begin by proving a new isoperimetric bound on the mixing time using a quantity which we call blocking conductance $\phi(x)$, this is an extension of conductance Φ and average conductance $\Phi(x)$. We then look at the three methods for bounding conductance of which we are aware : geometry, induction, and canonical paths. We extend all three of these methods and obtain bounds on the blocking conductance $\phi(x)$ or the conductance function $\Phi(x)$; in all three cases these give significant improvements over conductance based bounds for the mixing time. We end by considering a new isoperimetric quantity $h_2^+(x)$; we prove a mixing time bound in terms of $h_2^+(x)$ and conjecture a stronger theorem which may give optimal mixing time bounds for a wide range of Markov chains including geometric, inductive, and product Markov chains.

Chapter 1

Introduction

1.1 Mixing Times

This thesis is concerned with geometric methods for studying the rate at which Markov chains approach their steady state distributions.

Given a finite state, discrete time, Markov chain $\mathcal{M}$ – such as a random walk on a graph or a group – with state space Ω and transition probability matrix P, we are are interested in studying the rate at which the total variation distance, $\|\mathsf{p}^{(t)} - \pi\|_{TV}$ decreases to 0. In particular, given unknown initial distribution $\mathsf{p}^{(0)}$ and t-step distribution $\mathsf{p}^{(t)}$, we seek to find the smallest τ such that

$$\tau = \max_{\mathsf{p}^{(0)}} \min\{t : \|\mathsf{p}^{(t)} - \pi\|_{TV} \leq 1/4\} \ ;$$

this is generally known as the *mixing time*. It is known [ea] that for time-reversible Markov chains, to be defined later, we then have

$$\max_{\mathsf{p}^{(0)}} \|\mathsf{p}^{(\tau \log_2(1/\epsilon))} - \pi\|_{TV} \leq \epsilon \ ,$$

so upper bounding τ gives a first order estimate on the rate of convergence to the steady state distribution.

1.2 Applications

The recent interest in the study of the mixing time of Markov chains has been motivated in large part due to applications in computer science and physics which use the Markov Chain Monte Carlo method.

The *Markov Chain Monte Carlo* (MCMC) method is the use of Markov chains to sample from a fixed state space Ω and distribution σ, where the Markov Chain $\mathcal{M}$ is constructed so that its stationary distribution π equals σ. Then running the Markov chain for a large enough number of steps will give a sample from Ω that is very close to sampling by $\pi = \sigma$.

When the target distribution σ is uniform over Ω and there is a clear choice of a neighborhood structure (edges), then one way to generate σ is to set

$$\mathsf{P}(x,y) = \begin{cases} 1/d^* & if\ y \in \Gamma(x) \\ 1 - d(x)/d^* & if\ y = x \\ 0 & otherwise \end{cases} \tag{1.1}$$

where $\Gamma(x) = \{y \in \Omega : y \text{ is a neighbor of } x\}$, $d(x) = |\Gamma(x)|$ and $d^* \geq \max_{x \in X} d(x)$. The *Metropolis Method* can be used on top of this to obtain an arbitrary stationary distribution.

For example, let G be a graph and we want to sample uniformly from all k-colorings of G. Then the state space Ω is the space of all colorings of the vertices of G by k colors, where no two adjacent vertices have the same color; one reasonable neighborhood structure is to allow two colorings C_1 and C_2 to be adjacent if they differ at only a single vertex. The Markov chain $\mathcal{M}$ described above is then the *Glauber dynamics*, where steps are made by choosing a vertex $v \in G$ uniformly at random, a color c also uniformly, and then recoloring v to color c if this is possible (i.e. if no $y \in \Gamma(v)$ has color c). Then the stationary distribution of $\mathcal{M}$ is uniform over Ω.

There are three main applications for the Markov Chain Monte Carlo method.

- **Approximate Counting and Integration :** Approximate counting for self-reducible problems can be reduced to a uniform sampling problem [JVV86]. For example, the glauber dynamics above is commonly used to count the number of k-colorings of a graph G. Discrete integration (estimate $\sum_{x\in\Omega} w(x)$ for a positive weight

w) is a natural generalization, and is used in many volume algorithms [DFK91].

- **Statistical Physics :** Estimate the expectation of a random variable over a large complicated state space. For example, over arrangements of water molecules, distributions of atoms in a lattice, etc.

- **Combinatorial Optimization :** Construct $\mathcal{M}$ so that π is higher near optimal solutions and use this to search for optima. For example, simulated annealing.

Problems in approximate counting have proven the largest motivation of these three. It is a classical problem of combinatorics to estimate the size of a class of objects with a certain structure; for example, the number of k-colorings or Hamiltonian cycles of a graph. Complexity theory tells us that many counting problems are #P-hard – including the k-colorings and Hamiltonian cycle problems – and hence there is no efficient deterministic method (unless $P = NP$) for computing exactly the size of these objects. The next best solution is to look for approximation algorithms, and often the only known way to obtain good estimates for these counting problems are through randomized algorithms such as the Markov Chain Monte Carlo Method.

1.3 History of this field

Since the development of stochastic processes there has been much interest in convergence of Markov chains. An early question was whether a Markov chain was ergodic, that is whether it has a unique stationary distribution to which the Markov chain converges asymptotically. It was found that for the finite state discrete time Markov chains with which we are concerned, a Markov chain that is connected and aperiodic will asymptotically converge to a unique stationary distribution. Metropolis et.al. made use of this fact in their groundbreaking paper "Equation of state calculation by fast computing machines" [MRR$^+$53], where they constructed Markov chains with distributions applicable in Physics.

Exact information on the rate of convergence can be found from the spectrum of the transition probability matrix P. However, the spectrum is known for only a few highly symmetric Markov chains, such as a random walk around a cycle. An easier quantity to study is the *spectral gap* λ between the largest eigenvalue (one) and the second largest eigenvalue, as this is the dominant component of the mixing time [DS91]; when π_0 is the smallest steady state probability of a state then

$$\tau = O(\lambda^{-1} \log \pi_0^{-1}) \; .$$

It has proven difficult to obtain decent lower bounds on even the spectral gap, until recently.

One of the earlier successes in obtaining good upper bounds on the mixing time was the Coupling Theorem [Ald83], first used by Aldous. This led to simple accurate estimates on the mixing time for symmetric Markov chains such as random walks on the cycle, binary hypercube, and some other groups. Aldous and Diaconis used a related notion of the Strong Stopping Time [AD86] to study card shuffling. Applications include the famous result that seven shuffles is enough to shuffle a deck of cards. However, none of these methods seemed to apply to Markov chains much more complex than random walks on groups.

Jerrum and Sinclair [JS88] studied a geometric quantity they call the conductance Φ, a discretized version of Cheeger's constant h from Riemannian Geometry, and showed how it can be used to bound the spectral gap. When A is used to denote subsets of the state space Ω, and A^c denotes the complement $A^c = \Omega \setminus A$ then

$$\Phi = \min_{0<\pi(A)\leq 1/2} \frac{\sum_{\alpha \in A} \pi(\alpha)\, \mathsf{P}(\alpha, A^c)}{\pi(A)} \; .$$

They showed that $\frac{1}{2}\,\Phi^2 \leq \lambda \leq 2\Phi$, and so

$$\tau = O(\Phi^{-2} \log \pi_0^{-1}) \; .$$

The conductance of a Markov chain can be thought of as a measure of the smallest "ergodic flow" $\mathsf{Q}(A, A^c) = \sum_{\alpha \in A} \pi(\alpha)\, \mathsf{P}(\alpha, A^c)$ from a subset A to its complement A^c, relative to the size $\pi(A)$ of A; alternatively, it is a measure of the worst edge bottleneck in its underlying graph. Jerrum and Sinclair developed a practical method for lower bounding the conductance and used this method to bound the mixing time of a Markov chain of interest to computer scientists.

This led to a stream of new results and various methods to bound the conductance. In their paper Jerrum and Sinclair used a method known as canonical paths, Dyer, Frieze, and Kannan [DFK91] used isoperimetric inequalities to bound conductance for a Markov chain for approximating the volume of convex bodies, and Mihail and Sudan [MS92] used inductive methods for a Markov chain used to estimate the number of matroids with a given set of bases.

Two weaknesses with conductance are the large difference between the upper and lower bounds it gives on the spectral gap ($\frac{1}{2}\,\Phi^2 \leq \lambda \leq 2\,\Phi$), and the large difference between the upper and lower bounds that the spectral gap gives on the mixing time ($\lambda^{-1} \leq \tau \leq 2\lambda^{-1}\,\log \pi_0^{-1}$). Sinclair [Sin92] and Diaconis and Stroock [DS91] used canonical paths to bound the spectral gap directly, and thus gave tighter bounds on the spectral gap. Diaconis and Saloff-Coste [DSC96] extended ideas from the theory of hypercontractivity and showed how the *logarithmic-Sobolev constant* ρ, to be defined and discussed in Chapter 2 bounds mixing time ($\tau \leq 2\rho^{-1}\log\log \pi_0^{-1}$); ρ is often the same order as λ, in which case it gives tighter bounds on mixing time than does the spectral gap. Unfortunately, the log-Sobolev constant has proven difficult to work with except for with some very symmetric random walks on groups.

Bubley and Dyer [BD97] made the Coupling Theorem more practical with the development of *path coupling*. This was used by many authors [Wil97, Vig99] to obtain optimal results when a Markov chain has good local behavior. However, Kumar and Ramesh [KR01] showed that coupling and path coupling methods are unlikely to apply to several Markov chains of interest.

A new development in geometric methods was made by Lovász and Kannan [LK99], who developed the notion of the *conductance function*

$$\Phi(x) = \min_{0<\pi(A)\leq x} \frac{\sum_{\alpha\in A} \pi(\alpha)\,\mathsf{P}(\alpha,A^c)}{\pi(A)}\ ,$$

an extension of conductance that measures edge bottlenecks for sets of various sizes, and gave a proof that it can be used to bound the mixing time <u>directly</u>, without reference to the spectral gap,

$$\tau = O\left(\int_{\pi_0}^{1/2} \frac{dx}{x\,\Phi(x)^2} + \frac{1}{\Phi}\right)\ .$$

By using the conductance function it is possible to show optimal bounds on the mixing time of some simple Markov chains such a random walk on a cycle, whereas spectral gap or logarithmic-Sobolev methods are unable to do this. Lovász and Kannan gave an application of their method to a Markov chain of interest to them, however it is still an open problem to apply the conductance function to other Markov chains for which conductance has been used.

Independently, Houdré [Hou01] showed how to bound the log-Sobolev constant in terms of the spectral gap and geometric quantities closely related to the conductance function. However, he only gave applications to highly symmetric Markov chains.

1.4 The Results

In Chapter 3 we prove an extension of Average Conductance which we call Blocking Conductance. This is a method appearing in a paper by Kannan, Lovász and Montenegro [KLM02]. A key problem with conductance results has been that mixing time bounds always involve a square Φ^2 or $\Phi(x)^2$, so that we cannot expect to show the correct bounds when the mixing time is "square free" : for example, the lazy random walk on the binary hypercube $\{0,1\}^n$ has $\tau = \Theta(n \log n)$. With blocking conductance we will replace the $\Phi(x)^2$ with a term $\phi(x)$, thus eliminating the square and opening the possibility of better geometric mixing time bounds.

The blocking conductance function (BCF) $\phi(x)$ will involve a measure of edge×vertex isoperimetry, so that when $\phi(x)$ is high then this will mean that there are no sets of size x which simultaneously have an edge and a vertex bottleneck. This turns out to be a useful measure because in many problems the tightest edge bottlenecks only occur when there are many neighboring vertices, while vertex bottlenecks only occur when there are many edges. The binary hypercube $\{0,1\}^n$ is a classic example.

Next, we look for applications of average and blocking conductance. Both $\Phi(x)$ and $\phi(x)$ are extensions of Φ, so that we cannot expect to bound either quantity without knowing how to bound Φ. We are aware of three methods for bounding Φ – isoperimetry, induction and canonical paths – and by Jerrum and Sinclair's results (see Theorem 2.1) these give a method for bounding the mixing time. Most of Chapters 4-6 of this thesis will consist of extending these three methods in order to obtain bounds on $\Phi(x)$ or $\phi(x)$, and then by applying Theorem 3.1 or 3.2 we can significantly improve on previous geometric mixing time bounds in all three cases.

As the first application, in Chapter 4 we look at using isoperimetric inequalities to bound $\Phi(x)$. Recall that bounding the conductance Φ or conductance function $\Phi(x)$ is equivalent to finding a lower bound on the ratio of ergodic flow (edges) from sets A to A^c, relative to the size of A. Then the idea of isoperimetric inequalities is to consider Markov chains whose underlying graph can be embedded in $\mathbb{R}^n$, while preserving adjacency (edges) and set size (size of A). For example, the n-dimensional grid $[k]^n$ can be embedded by associating points $v = (v_1, v_2, \ldots, v_n)$ with cubes $P(v) = \{x = (x_1, x_2, \ldots, x_n) : v_i - \frac{1}{2} \leq x_i \leq v_i + \frac{1}{2}\}$. If A is a subset of the state space Ω, then such an embedding will be such that the volume $vol_n P(A)$ is proportional to $\pi(A)$ and the surface area $vol_{n-1} \partial P(A)$ is proportional to the number of edges (or flow) from A to A^c.

With this in mind, it suffices to find an isoperimetric inequality for convex bodies $K \subset \mathbb{R}^n$, with a log-concave distribution F on K, and subsets $S \subset K$ with piecewise smooth boundaries; the inequality should give a lower bound on the ratio of surface area $\int_{\partial S \setminus \partial K} F(y)\, dy$ to the volume $\int_S F(y)\, dy$. This is known to suffice for bounding the conductance and conductance function of certain Markov chains [KK91, DF91, LS93]. Past results have focused on finding the worst such ratio among all subsets covering at most half of the space K (i.e. $\int_S F(y)\, dy \leq \frac{1}{2} \int_K F(y)\, dy$), but we will need to bound the ratio among all subsets of a fixed size x (i.e. $\int_S F(y)\, dy = x \int_K F(y)\, dy$).

Theorem 4.1 gives such a lower bound and has the nice property that it is tight for every set size x. In particular, if the space is the unit hypercube $K = [0,1]^n$ and x is fixed, then we exhibit a log-concave distribution F on K and a subset A such that the lower bound is an equality. Moreover, when the distribution is restricted to uniform ($F = 1$) then we derive a slightly stronger result, Theorem 4.7, which is again tight for every x and dimension n.

Theorem 4.1 can be used to obtain improved bounds for mixing time on the Markov chains to which isoperimetric inequalities have been applied [KK91, DF91, LS93]. We are able to show that when previous methods gave conductance Φ_g and mixing time $\tau = O(\Phi_g^{-2} \log \pi_0^{-1})$, our new isoperimetric inequality on $\Phi(x)$ gives mixing time $\tau = O(\Phi_g^{-2})$, thus saving a factor of $\log \pi_0^{-1}$. We also give the first non-trivial bounds on the log-Sobolev constants of these geometric Markov chains.

This can give a substantial improvement. For the lazy random walk on the grid $[k]^n$ the improvement is from $\tau = O(n^3 k^2 \log k)$ to $\tau = O(n^2 k^2)$, while the optimal bound is $\tau = \Theta(k^2 n \log n)$. Example 4.8 deals with a Markov chain on linear extensions, this improves from $\tau = O(n^5 \log n)$ [Jer98] to $\tau = O(n^4)$ and even beats the path-coupling and comparison bound [BD97] of $\tau = O(n^4 \log^2 n)$. Wilson [Wil97] used an elegant path coupling to show the correct bound is $\tau = \Theta(n^3 \log n)$, so our results are quite close to optimal. (We note that in her Ph.D. dissertation Chen [Che00] gave a different method of sampling linear extensions which sometimes mixes in time $O(n^2 \log n)$).

Next, in Chapter 5 we look at an inductive method of bounding conductance. The best example of which we are aware is the use of induction [MS92, FM92] to bound the conductance of a Markov chain on matroids with a given sets of bases. Matroids are a generalization of the linear algebraic notions of linear independence and bases, which appear in different guises in many areas of mathematics. A matroid can be thought of as having a set of elements $E(\mathcal{M})$ of size $m = |E(\mathcal{M})|$ and a collection of bases $\mathcal{B}(\mathcal{M})$ each of size n (the rank). The aforementioned papers considered a class of matroids known as balanced matroids, and which have a natural inductive structure.

We are able to extend their inductive methods and obtain a bound on $\Phi(x)$. In Corollary 5.2 we apply Houdré's Theorem 2.9 to this $\Phi(x)$ and bound the log-Sobolev constant for this Markov chain, and obtain a mixing time of $\tau = O(m^{3/2} n^2 \log\log(m^n))$, a significant improvement over the conductance based bound [FM92, MS92] of $\tau = O(m^2 n^3 \log m)$. For regular matroids with a constant number of parallel edges, such as graphic matroids without multiple edges, this is also an improvement over Feder and Mihail's canonical path bound of $\tau = O(m n^3 \log m)$ [FM92]. It is worth noting that this is one of the more complicated problems studied with log-Sobolev constants.

In Chapter 6 look at applying blocking conductance $\phi(x)$. As a first application we consider canonical paths methods, the third and most common method of bounding conductance. As mentioned in the previous section, canonical paths can be used to bound the spectral gap directly [Sin92], and so conductance is generally not used in this case. Nevertheless, we are able to use a geometric argument based on blocking conductance to obtain Theorem 6.1, a mixing time bound closely related to Sinclair's spectral gap results (see Theorem 2.6).

Many results previously shown by applying canonical paths to the spectral gap can be shown equally well with Theorem 6.1. For example, Feder and Mihail's [FM92] canonical path bound on mixing time of balanced matroids carries over exactly. Then, between Corollary 5.2 and Theorem 6.1 our isoperimetric results are always at least as strong, and sometimes stronger than their spectral gap bounds.

The main result in Chapter 6 uses blocking conductance to show a new type of isoperimetric bound on mixing time. We consider the quantity $h_2^+(x)$, which is a member of the family

$$h_p^+(x) = \min_{\substack{A \subset \Omega \\ 0 < \pi(A) \leq x}} \frac{\sum_{\alpha \in A} \pi(\alpha) \sqrt[p]{\mathsf{P}(\alpha, A^c)}}{\pi(A)\,\pi(A^c)} \ .$$

It is clear that $\frac{1}{2}\, h_1^+(x) \leq \Phi(x) \leq h_1^+(x)$, and because $\mathsf{P}(\alpha, A^c) \leq 1$ then $h_2^+(x)$ can be substantially larger than $\Phi(x)$.

A restricted version $h_p^+ = h_p^+(1/2)$ has been considered by Bobkov, Houdré and Tetali [BHT00, HT96] (see Example 6.3 in this thesis), where $h = h_1^+$ is twice the Conductance / Cheeger constant. It is known that h_1^+ relates to edge isoperimetry and that h_∞^+ relates to vertex isoperimetry; an application of Cauchy-Schwarz shows that h_2^+ relates to a mixture of edge and vertex isoperimetry [Tal93]. h_1^+ and h_∞^+ can both be used to bound the spectral gap λ [BHT00], and thus the mixing time as well. In Theorem 6.3 we are able to upper and lower bound the quantity $\Psi_{int}(A)$, a form of blocking conductance defined on sets, in terms of $h_2^+(A)$, the value of h_2^+ on the set A. We obtain

$$2\, h_2^+(A)^2 \geq \Psi_{int}(A) \geq \frac{h_2^+(A)^2}{(2 + \log(1/\sqrt{\mathsf{P}_{min}}))^2} \ ,$$

where P_{min} is the smallest non-zero transition probability. This leads to a better mixing time bound than that available by either h_1^+ or h_∞^+. We get that

$$\tau = O\left(\log^2(1/\mathsf{P}_{min}) \int_{\pi_0}^{1/2} \frac{dx}{x\, h_2^+(x)^2} + \frac{1}{\Phi} \right) \ .$$

This bound is roughly the Average Conductance theorem with $h_2^+(x)$ replacing $\Phi(x)$, and with the (small) extra $\log^2(1/\mathsf{P}_{min})$ term, and so it may give a substantial improvement over bounds involving $\Phi(x)$.

There are several applications of this. Talagrand [Tal93] has bounded $h_2^+(x)$ for the binary hypercube $\{0,1\}^n$. When substituted into Theorem 6.3 this shows that $\tau = O(n \log^3 n)$, which is a significant improvement over the Average Conductance bound of $\tau = O(n^2)$ (see Chapter 4). In Example 6.3 we apply results of [HT96] to bound $h_2^+(x)$ for product Markov chains, this gives a mixing times bound which is almost optimal for products. Example 6.4 deals with a related quantity $\tilde{\beta}_2$ defined in [Mur01], by using Theorem 6.3 we show that $\tilde{\beta}_2^2$ is nearly as strong as log-Sobolev when dealing with mixing time.

In Chapter 7 it is conjectured that the extra terms in Theorem 6.3 are unnecessary, so that the mixing time is bounded by

$$\tau = O\left(\int_{\pi_0}^{1/2+\pi_{max}} \frac{dx}{x\, h_2^+(x)^2}\right) ,$$

where $\pi_{max} = \max \pi(x)$ is the maximal stationary distribution of a point. If this is the case then the bound on the binary hypercube and on general product Markov chains will be tight. Moreover it would also show that $\tilde{\beta}_2^2$ is as strong as the log-Sobolev constant for showing mixing time; it may turn out that $\tilde{\beta}_2^2$ is a much easier quantity to bound, because it has no logarithmic terms and bears a strong resemblance to the conductance Φ. It also seems likely that $h_2^+(x)$ will give tight bounds for both the Markov chain on linear extensions and the one on balanced matroids.

Chapter 2

Preliminaries

2.1 Markov Chains

In this thesis we are concerned with studying finite state Markov chains $\mathcal{M}$. Finite state Markov chains can be interpreted in many ways, we will consider them either as a random walk on a weighted graph (E, V) or as a product of stochastic matrices P.

A *finite state Markov chain* $\mathcal{M}$ is given by a state space Ω with cardinality $|\Omega| = n$, and the *transition probability matrix*, an $n \times n$ square matrix P such that $\mathsf{P}_{ij} \in [0,1]$ and

$$\forall i \in \Omega \;:\; \sum_{j \in \Omega} \mathsf{P}_{ij} = 1 \;.$$

A Markov chain is *connected* if $\forall i, j \in \Omega,\ \exists t : \mathsf{P}^t_{ij} > 0$. Moreover, it is *aperiodic* if $\forall i : gcd\{t : (\mathsf{P}^t)_{ii} > 0\} = 1$. This guarantees that $\exists N : t \geq N \implies (\mathsf{P}^t)_{ii} > 0$, i.e. after enough steps of the Markov chain there will always be a non-zero probability of being at any point i. A Markov chain is *lazy* if $\forall i \in \Omega : \mathsf{P}_{ii} \geq \frac{1}{2}$; lazy chains are obviously aperiodic.

The *initial distribution* on Ω is given by an n-dimensional column matrix $\mathsf{p}^{(0)}$, such that $(\mathsf{p}^{(0)})_i \in [0,1]$ and $\sum_{i \in V} (\mathsf{p}^{(0)})_i = 1$. The transition probability matrix P then determines the distribution at later times, we write $\mathsf{p}^{(t)}$ to denote the distribution after t steps of the Markov chain $\mathsf{p}^{(t)}(i) = \sum_{j \in \Omega} \mathsf{p}^{(t-1)}(j)\, \mathsf{P}_{ji}$.

A key quantity with which we are concerned is the *stationary distribution* (or *steady-state distribution*) π of a Markov chain $\mathcal{M}$. This is a distribution π on Ω such that $\pi\, \mathsf{P} = \pi$. It is well known that a connected aperiodic finite state discrete time Markov chain has exactly one stationary distribution π.

A Markov chain with stationary distribution π is *ergodic* if $\forall i,\, j \in \Omega : \lim_{t\to\infty} \mathsf{P}^t_{ij} = \pi_j$. In particular, connected aperiodic finite state discrete time Markov chains are ergodic. This thesis will study how quickly ergodic Markov chains approach their stationary distribution.

The *flow* between two points $i, j \in \Omega$ is $q(i,j) = \pi_i \, \mathsf{P}_{ij}$ – the $q(i,j)$ can be thought of as edge weights for the directed graph $\Omega \times \Omega$ – and the flow between two sets $A,\, C \subset \Omega$ is

$$Q(A,C) = \sum_{\substack{i\in A\\ j\in C}} q(i,j) = \sum_{\substack{i\in A\\ j\in C}} \pi_i \, \mathsf{P}_{ij} \ .$$

The Markov chains we study will be *time-reversible*, that is $\forall i, j \in \Omega : \pi_i \, \mathsf{P}_{ij} = \pi_j \, \mathsf{P}_{ji}$, or equivalently $q(i,j) = q(j,i)$ so the graph $\Omega \times \Omega$ is undirected. This says that the flow along an edge is the same in both directions, so in particular an outside observer watching vertices and edges of a Markov chain in the steady state would not be able to distinguish between a step of the Markov chain P and the reverse time Markov chain P^{-1}.

A nice fact about Markov chains is that if a distribution σ satisfies $\sigma_i \, \mathsf{P}_{ij} = \sigma_j \, \mathsf{P}_{ji}$, then it follows that σ is the stationary distribution and the Markov chain is time-reversible. For example, if P is *symmetric* (i.e. $\forall i, j \in \Omega : \mathsf{P}_{ij} = \mathsf{P}_{ji}$) and $\sigma = 1/|\Omega|$ is uniform then trivially $\sigma_i \, \mathsf{P}_{ij} = \sigma_j \, \mathsf{P}_{ji}$; it follows that symmetric Markov chains have uniform stationary distribution. Another easy example to check is that the Markov chain given by (1.1) in the introduction has uniform stationary distribution.

Unless otherwise stated, all Markov chains in this thesis will be assumed to be ergodic, lazy, time-reversible, finite state discrete time Markov chains. These are completely natural assumptions; the laziness eliminates all aperiodic behavior while increasing mixing time by only a factor of 2, time-reversibility allows for simple computation of the stationary distribution and also allows the problem to be considered as a weighted undirected graph, and combinatorial problems have a finite number of states.

2.2 Various Bounds on the Mixing Time

2.2.1 The Mixing Time

For probability distributions σ and τ on Ω the *total variation distance* (often just called *variation distance*) is

$$\|\sigma - \tau\|_{TV} = \frac{1}{2}\sum_{i\in\Omega} |\sigma(i) - \tau(i)| = \max_{A\subset\Omega}(\sigma(A) - \tau(A)) \ .$$

This is a measure of how far σ and τ are from equality.

The *mixing time* $\tau(\epsilon)$ measures how many steps it takes a Markov chain to approach the stationary distribution from the worst starting distribution. It is not difficult to see that for finite state spaces, the worst initial distribution will be to start at a single point. Then we let

$$\tau(\epsilon) = \max_{i\in\Omega} \min \left\{t : \|\delta_i \, \mathsf{P}^t - \pi\|_{TV} \leq \epsilon\right\}$$

where δ_i is the distribution with $\delta_i(j) = 1$ if $j = i$ and 0 otherwise.

By convention we will define the *mixing time* to be $\tau = \tau(1/4)$. This suffices to bound $\tau(\epsilon)$ because [AF],

$$\|\delta_i \mathsf{P}^{\tau \log_2(1/\epsilon)} - \pi\|_{TV} \leq \epsilon \ ,$$

i.e. $\tau(\epsilon) \leq \tau \, \log_2(1/\epsilon)$.

2.2.2 Equivalence of Different Bounds

An important concept when studying mixing times is the equivalence of various types of mixing times. In almost all cases, methods for bounding the mixing time do so by bounding some other quantity, such as spectral gap λ, which is known in turn to bound the total variation distance.

A few examples of equivalences that we will use are as follows. These equivalences are the results of work by Aldous, Lovász and Winkler [ALW97, LW96]. First we need a few definitions. In all cases we will be working with lazy, time-reversible, ergodic Markov chains $\mathcal{M}$ with stationary distribution π.

Given a Markov chain $\mathcal{M}$, initial distribution μ and target distribution σ, a *stopping rule* is a rule which observes the progress of the Markov chain and stops at some randomized time Λ, such that the final distribution is σ. The simplest stopping rule would choose a state $x \in \Omega$ according to σ, and then let the Markov chain proceed, stopping when it reaches x. Define

$$h(\mu, \sigma) = \inf\{\mathsf{E}_\mu T : T \text{ is a randomized stopping time from } \mu \text{ to } \sigma\} .$$

Then the *forget time* is

$$\mathcal{T}_{forget} = \inf_\sigma \sup_\mu h(\mu, \sigma) = \inf_\sigma \sup_x h(x, \sigma) .$$

This is the time to reach the distribution σ which is closest to all initial distributions; once σ has been reached then the Markov chain has forgotten all information about the starting distribution.

The *reset time*

$$\mathcal{T}_{reset} = \sum_j \pi_j \, h(j, \pi)$$

is the expected time to reach π when the initial distribution is drawn from π.

The *mixing time* $\mathcal{T}_{mix}$,

$$\mathcal{T}_{mix} = \sup_\mu h(\mu, \pi) ,$$

is the expected time to reach π with an optimal stopping rule and the worst initial distribution. The stopping rule in this case is often called *mean optimal*.

Let $\rho_t(\cdot) = \frac{1}{t} \sum_{i=0}^{t-1} \mathsf{p}^{(i)}(\cdot)$ be the t-step average distribution of the Markov chain. Then

$$\mathcal{T}_{uniform} = \min\{t : \|\rho_t - \pi\|_{TV} \leq \frac{1}{4}\}$$

is the time for the average distribution to approach stationary.

Then we have :

- $\mathcal{T}_{forget} \leq \mathcal{T}_{mix}$ (obvious).
- $\mathcal{T}_{reset} \leq \mathcal{T}_{mix}$ (obvious).
- $\mathcal{T}_{mix} \leq 2(\mathcal{T}_{reset} + \mathcal{T}_{forget})$. [ALW97]
- $\mathcal{T}_{forget} \leq 43\, \mathcal{T}_{uniform}(1/4)$. [ALW97]

- If $\mathcal{M}$ is time-reversible then $T_{forget} = \mathcal{T}_{reset}$. [LW96]
- If $\mathcal{M}$ is time-reversible then $\tau(\epsilon) \leq 8\,\mathcal{T}_{mix}\,\log_2(1/\epsilon)$. [ea]
- If $\mathcal{M}$ is time-reversible then $\tau(\epsilon) \leq 1376\,\mathcal{T}_{uniform}(1/4)\,\log_2(1/\epsilon)$.

The proof of the Average Conductance theorem (Theorem 2.2) bounds $\mathcal{T}_{mix}$, while the theorem (Theorem 2.3) using the Bounded Conductance Function will bound $\mathcal{T}_{uniform}$. Clearly the constant 1376 used to convert between $\mathcal{T}_{uniform}(1/4)$ and $\tau(\epsilon)$ is impractical, so it is to be hoped that a direct proof can improve on this. However, we are interested mainly in the order of magnitude of the mixing time in terms of some measure n of the size of the problem, and not so much in the constant term, so we often state results in terms of big-O $O(\cdot)$ to ignore the conversion between mixing times.

- $f(n) = O(g(n))$ if $\exists C \in \mathbb{R}^+ : \forall n,\ f(n) \leq C\,g(n)$.
- $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$, or equivalently if $\exists c \in \mathbb{R}^+ : \forall n,\ f(n) \geq c\,g(n)$.
- $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$, or equivalently if $\exists c, C \in \mathbb{R}^+ : \forall n,\ c\,g(n) \leq f(n) \leq C\,g(n)$.

2.2.3 Isoperimetric Bounds

A common geometric technique for bounding the mixing time on complicated Markov chains involves *conductance*, or more generally the *conductance function*. The *conductance function* of $\mathcal{M}$ is defined in terms of the flow,

$$\Phi(x) = \min_{\substack{S\subset\Omega \\ \pi_0\leq\pi(S)\leq x}} \frac{Q(S,S^c)}{\pi(S)}$$

where $\pi_0 = \min_{v\in V} \pi_v$ and $\pi_0 \leq x \leq 1/2$. Also, the *conductance* Φ is defined by $\Phi = \Phi(1/2)$.

If a finite state Markov chain has a uniform (i.e. constant) stationary distribution and all transition probabilities are either a constant p or 0, then bounding the conductance function is equivalent to bounding the *edge-isoperimetry*, because

$$\Phi(x) = \min_{\pi_0\leq\pi(S)\leq x} \frac{Q(S,S^c)}{\pi(S)} = p \min_{0<|S|\leq x|\Omega|} \frac{|Cut(S)|}{|S|}\ , \tag{2.1}$$

where $Cut(S) = \{(i,j) \in S \times S^c : P_{ij} > 0\}$. Likewise, bounding the conductance is equivalent to bounding *cutset expansion.* More generally, as long as the Markov chain is time-reversible then the conductance can be reduced to the edge-isoperimetry of a weighted graph. This motivates the use of edge-isoperimetry as a means for bounding the conductance function.

The following two theorems can be used to bound the mixing time.

Theorem 2.1 (Conductance [JS88]). *The mixing time τ of any Markov chain is bounded by*

$$\tau \leq \frac{2}{\Phi^2} \log(4/\pi_0)) \ .$$

Theorem 2.2 (Average Conductance [LK99]). *The mixing time τ of any Markov chain is bounded by*

$$\tau \leq K \left(\frac{4}{\Phi} + 14 \int_{\pi_0}^{1/2} \frac{dx}{x\Phi(x)^2} \right) \ ,$$

where $K = 16$ arises from converting between $\mathcal{T}_{mix}$ and τ.

Theorem 2.2 is essentially that given in [LK99], however we have corrected a minor mistake in their theorem (the $4/\Phi$ term was omitted) and have adjusted the constants to take into account the fact that our conductance function differs from theirs by roughly a factor of 2. The corrections and a proof of Theorem 2.2 can be found in Chapter 3.

We will also show rapid mixing by a newer method, using a quantity known as a blocking conductance function $\phi(x)$ (BCF). We refer the reader to Definition 3.2 in Chapter 3 for the definition of a BCF $\phi(x)$, but we note that roughly speaking $\phi(x) \geq \frac{1}{4}\Phi^2(x)$ so the following theorem promises to be at least as strong as Theorem 2.2.

Theorem 2.3. *If $\mathcal{M}$ is a Markov chain and $\phi(\cdot)$ is a blocking conductance function, then*

$$\tau \leq K \int_{\pi_0}^{1/2} \frac{dx}{x\,\phi(x)} \ ,$$

where K is a constant independent of the Markov chain.

We are aware of three methods for bounding Φ, and by Theorem 2.1 these give a method for bounding the mixing time. We will extend all of these methods in order to obtain bounds on $\Phi(x)$ or $\phi(x)$, then by applying Theorem 2.2 or Theorem 2.3 we will significantly improve on previous mixing time results in all three cases.

2.2.4 Spectral gap and log-Sobolev

Many proofs of rapid mixing use the *spectral gap* λ or the *log-Sobolev constant* ρ. The spectral gap λ is the difference between the largest eigenvalue of P (i.e. 1, with eigenvector π) and the second largest eigenvalue. The log-Sobolev constant is a sort of local entropy of the Markov chain and arises in the theory of hypercontractivity.

$$\lambda = \inf_{\substack{f:\Omega\to\mathbb{R} \\ Var(f)\neq 0}} \frac{\mathcal{E}(f,f)}{Var(f)} \qquad \text{and} \qquad \rho = \inf_{\substack{f:\Omega\to\mathbb{R} \\ \mathcal{L}(f)\neq 0}} \frac{\mathcal{E}(f,f)}{\mathcal{L}(f)}$$

where $\mathrm{Var}(f)$ is the variance of f, $\mathcal{E}$ is the Dirichlet form, and $\mathcal{L}$ is the entropy

$$\begin{aligned} \mathcal{E}(f,f) &= \frac{1}{2}\sum_{x,y\in\Omega} (f(y)-f(x))^2\pi(x)\,P(x,y) \\ Var(f) &= \frac{1}{2}\sum_{x,y\in\Omega} (f(y)-f(x))^2\pi(y)\pi(x) \\ \mathcal{L}(f) &= \sum_{x\in\Omega} |f(x)|^2\log\left(\frac{|f(x)|^2}{\|f\|_2}\right)\pi(x) \end{aligned}$$

The spectral gap is well known to bound the mixing time, because the second eigenvalue is the slowest component to go to zero. In particular, we have the following theorem [DS91].

Theorem 2.4. *The mixing time τ of a Markov chain $\mathcal{M}$ can be bounded by*

$$\tau \le \frac{1}{\lambda}\,(2+\log\pi_0^{-1})$$

where λ is the spectral gap of $\mathcal{M}$.

We can derive Theorem 2.1 from this by using the following theorem [JS88].

Theorem 2.5. *The spectral gap of a Markov chain $\mathcal{M}$ satisfies*

$$\frac{\Phi^2}{2} \le \lambda \le 2\,\Phi$$

where Φ is the conductance of $\mathcal{M}$.

Perhaps the most common method used to bound the conductance or the mixing time is by canonical paths. For every pair of vertices x, $y \in \Omega$, let γ_{xy} be a path from x to y along edges in the underlying graph, and let e be used to denote edges in the graph. Ideally the paths should be chosen so that they are not heavily concentrated on any particular edge. Then Sinclair [Sin92] showed that

Theorem 2.6. *Let $\mathcal{M}$ be a Markov chain with underlying graph $G = (E, V)$ and let*

$$\rho_e = \max_{e \in E} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) \qquad \text{and} \qquad \overline{\rho} = \max_{e \in E} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)\, |\gamma_{xy}|$$

then

$$\Phi \geq \frac{1}{2\,\rho_e} \qquad \lambda \geq \frac{1}{\overline{\rho}} \geq \frac{1}{\rho_e\,\ell_{max}}$$
$$\tau \leq \rho_e\,\ell_{max}\,(2 + \log \pi_0^{-1})$$

where $\ell_{max} = \max_{x,y \in V} |\gamma_{xy}|$ denotes the length of the longest path.

In Chapter 6 we will show a closely related result by using only isoperimetry.

Theorem 2.7. *If $\mathcal{M}$ is a Markov chain with underlying graph $G = (E, V)$, define*

$$\rho_v = \max_{v \in V} \frac{1}{\pi(v)} \sum_{\gamma_{xy} \ni v} \pi(x)\,\pi(y) \quad \text{and} \quad \rho_v^{ave} = \sum_{v \in V} \pi(v) \left[\frac{1}{\pi(v)} \sum_{\gamma_{xy} \ni v} \pi(x)\,\pi(y) \right]$$

to be the maximal and average vertex congestion over the space.

Then

$$\begin{aligned} \tau &= O\left(\rho_v\,\rho_e\,\log \pi_0^{-1}\right) \\ &= O\left(\frac{\rho_v}{\rho_v^{ave}}\,\rho_e\,\ell_{ave}\,\log \pi_0^{-1}\right) , \end{aligned}$$

where $\ell_{ave} = \sum_{x,y \in V} \pi(x)\,\pi(y)\,|\gamma_{xy}|$ is the average length of the canonical paths.

Other methods, such as isoperimetric inequalities or inductive methods, have also been used to bound the conductance and will be discussed further in Chapters 4 and 5.

Some examples of log-Sobolev constants can be found in [DSC96]. In particular it is shown that

Theorem 2.8. *The mixing time τ of a Markov chain $\mathcal{M}$ can be bounded by*

$$\tau \leq \frac{1}{2\rho}(2 + \log\log \pi_0^{-1})$$

where ρ is the log-Sobolev constant of $\mathcal{M}$.

In theory, the log-Sobolev constant may be the same order as the spectral gap, and so Theorem 2.8 may improve spectral gap results. However, in practice the log-Sobolev constant has proven very difficult to compute and thus has been used for very few problems.

Recent work of Houdré [Hou01] gives reasonably good bounds on the log-Sobolev constant in terms of a quantity related to the conductance function. We rewrite the definitions and theorems from [Hou01] in a form which is equivalent, but where the relation to our current techniques is clearer.

Let

$$g_1^+ = \inf_{\pi_0 \le x \le 1/2} \frac{\Phi(x)}{\sqrt{\log(1/x)}} \qquad \text{and} \qquad \ell_1^+ = \inf_{\pi_0 \le x \le 1/2} \frac{\Phi(x)}{\log(1/x)} \ . \tag{2.2}$$

These quantities may seem a bit artificial, however it is shown in Section 3.1.3 that when combined with Theorem 2.2 the g_1^+ and ℓ_1^+ are natural improvements on Φ.

Then we have

Theorem 2.9 (Houdré). *Let $\mathcal{M}$ be a Markov chain and let g_1^+ and ℓ_1^+ be as in (2.2). Then the log-Sobolev constant ρ is bounded by*

$$(i)\ \rho \ge \frac{1}{100}(g_1^+)^2 \qquad (ii)\ \rho \ge \frac{\lambda \ell_1^+}{2(\sqrt{\lambda} + 2\ell_1^+)} \ge \frac{\sqrt{\lambda}\ell_1^+}{12} \ ,$$

where we simplified the final inequality by using the fact $2\lambda \ge \Phi^2 \ge (\ell_1^+ \log 2)^2$.

By combining Theorems 2.8 and 2.9 (i), we can obtain an improvement over the Conductance bound in Theorem 2.1 when $g_1^+ \approx \Phi$. From our simplification of Theorem 2.9 (ii), we see that the second log-Sobolev bound will improve on the spectral gap bound when $\lambda \approx \Phi^2$ and $\ell_1^+ \approx \Phi$.

There are other methods for bounding the mixing time, such as coupling or strong stopping times, however these do not relate directly to the methods considered in this paper and hence we do not discuss them here.

Chapter 3

Isoperimetric Bounds on Mixing Time

This chapter is concerned with establishing and proving results on the mixing time of Markov chains in terms of certain isoperimetric quantities.

The first method, known as *Average Conductance* uses a generalization of the conductance Φ known as the *conductance function* $\Phi(x)$, a measure of bottlenecks which depends on the set size x (see Section 2.2.3). The main theorem was first proved in [LK99].

The second method uses a quantity known as a *blocking conductance function* (BCF). This can be thought of as a generalization of $\Phi(x)^2$, which simultaneously considers both edge and vertex bottlenecks. The theorem is based on unpublished work of Kannan, Lovász and Montenegro [KLM02].

3.1 Average Conductance

3.1.1 The Theorem

The theorem as stated in [LK99] has an error, we give the correct form of the theorem below. The constants are different than in [LK99] because our definition of conductance differs from theirs by roughly a factor of 2.

Theorem 3.1 (Average Conductance [LK99]). *The mixing time τ of any Markov chain* *is bounded by*

$$\tau \leq K \left(14 \int_{\pi_0}^{1/2} \frac{dx}{x\Phi(x)^2} + \frac{4}{\Phi} \right)$$

where $K = 16$ arises from converting between different measures of mixing time.

The proof can be found in [LK99]. An alternative proof with different constants is given at the end of Section 3.2.2.

3.1.2 A Correction to Lovász and Kannan's Paper

Theorem 3.1 as given in [LK99] was incorrect as stated, because it left out the $4/\Phi$ term. A counterexample to the original form is given as the final example of Section 3.1.3.

The problem in the proof (see [LK99]) is as follows. Sets were constructed with sizes

$$\pi_0 = x_0 < x_1 < \ldots x_{n-1} \leq .5 < x_n \ .$$

However, the integral that was used estimated the terms up to x_n, but the remainder from 0.5 to x_n was ignored. There's a similar problem for the reverse direction going from $1 - \pi_0$ down to 0.5.

To compute the amount that was left out, look at the construction in the proof and observe that the integral left out was used to approximate the term $2/\Phi(x_{n-1})$. But, $2/\Phi(x_{n-1}) \leq 2/\Phi$, so it suffices to add a $2/\Phi$ term for the construction from π_0 to 0.5, and another $2/\Phi$ for the construction from $1 - \pi_0$ down to 0.5.

3.1.3 Some Examples of Average Conductance

Much of this thesis is concerned with bounding the conductance function, via edge-isoperimetry, for problems of combinatorial interest. Edge-isoperimetry is a well studied subject, some information on this topic and a list of references can be found in [Bez99].

The quantities g_1^+ and ℓ_1^+ defined in (2.2) can be combined with Theorem 3.1 to get some intriguing mixing time results. From the definitions we have that $\Phi(x) \geq g_1^+ \sqrt{\log(1/x)}$ and $\Phi(x) \geq \ell_1^+ \log(1/x)$. Substituting these bounds into the Average Conductance Theorem 3.1

and integrating gives

$$\tau \leq C \begin{cases} \frac{1}{\Phi^2} \log \pi_0^{-1} \\ \frac{1}{g_1^{+2}} \log\log \pi_0^{-1} \\ \frac{1}{\ell_1^{+2}} \end{cases} \tag{3.1}$$

with the constant C a constant independent of the Markov chain.

The first bound is the Jerrum and Sinclair's Theorem 2.1. The second term is just the mixing time bound given by combining Theorem 2.9 (1) and Theorem 2.8, so Average Conductance is always at least as good as the first part of Houdré's result. The third bound is entirely new. It is worth noting that the tensorization property of the log-Sobolev constant [Hou01] makes Theorem 2.9(1) more useful than Average Conductance when working with product chains.

The bounds in (3.1) show that g_1^+ and ℓ_1^+ are in a sense natural analogs of conductance. The different bounds are best under different circumstances. The problems in the next two chapters will be cases of $\ell_1^+ = \Omega(\Phi)$, in this case the third bound is the best. The first bound is best for the random walk on the complete graph K_n, we don't know of a situation when the second bound is best.

However, Theorem 3.1 says more than this simple generalization. Consider the random walk on the barbell given by two complete graphs K_n with transition probabilities $1/2n$ inside each K_n, connected by a single central edge, and with transition probability along the central edge of $\epsilon/2n$. Then $\Phi(x) \geq \max\{\frac{1}{2} - x, \frac{1}{2n}\}$ for $x < 1/2$, $\Phi(1/2) = \Theta(\epsilon/n^2)$ and Theorem 2.2 gives $\tau = O(n^2/\epsilon) = O(1/\Phi)$, the correct bound.

This shows that in special cases Average Conductance can even hit the lower bound in $1/\Phi \leq \tau \leq (2/\Phi^2)(2 + \log(1/\pi_0))$. It is also an example of why the correction to the original theorem is needed; the trouble occurs because $\Phi(x)$ does not depend on ϵ except at $x = 0.5$ (i.e. Φ), a set of measure 0 in the integral, so the original theorem doesn't take the ϵ into account.

3.2 Blocking Conductance Function

In this section we develop a stronger version of Average Conductance, which uses a quantity called the *Blocking Conductance Function* $\phi(x)$. This method uses a measure of *edge* $\times$ *vertex* isoperimetry, rather than the edge isoperimetry which was used by conductance or average conductance methods.

3.2.1 Definitions and the Theorem

As usual, let $\mathsf{p}^{(t)}$ denote the probability distribution after t steps of the Markov chain, so $\mathsf{p}^{(0)}$ is the initial distribution. Also, for $x \in \Omega$, where Ω is the state space, let

$$\rho^t(x) = \frac{1}{t} \sum_{i=0}^{t-1} \mathsf{p}^{(i)}(x)$$

be the average probability distribution. As mentioned in the preliminaries, in order to bound the mixing time it suffices to bound the time for ρ^t to approach stationary.

For $t \in \mathbb{Z}^+$ and $x \in [0,1]$, define

$$h^t(x) = \sup_{A:\pi(A)<x} \left(\mathsf{p}^{(t)}(A) - \pi(A) \right) \quad \text{and} \quad h^t(x) = 0 \text{ for } x < \pi_0, \ ,$$

and let $h(x) = h^0(x)$. It was shown in [LS90] that for fixed x, $h^t(x)$ is non-increasing with t, so $h^t(x) \leq h(x)$ and in particular if $A \subseteq \Omega$ then $\rho^t(A) - \pi(A) \leq h(\pi(A))$. Given $x_0 \in [0,1]$ we will assume that the initial distribution is not too overweighted on sets smaller than x_0, e.g. if $x_0 = \pi_0$ then any initial distribution will be allowed.

To motivate the definition of blocking conductance we start with a restricted form known as the *exterior blocking conductance function*. The idea of blocking conductance is to improve on the $\Phi(x) \cdot \Phi(x)$ in Theorem 3.1 (Average Conductance) by replacing the second $\Phi(x)$ with a quantity involving boundary vertices.

To do this, consider a set $A \subset \Omega$ and let $B \subset A^c$ with $\pi(B) = \frac{1}{2}\,\mathsf{Q}(A, A^c)$. Then

$$\mathsf{Q}(A,\, A^c \setminus B) = \mathsf{Q}(A, A^c) - \mathsf{Q}(A, B) \geq \mathsf{Q}(A, A^c) - \pi(B) \geq \frac{1}{2}\,\mathsf{Q}(A, A^c)\ , \tag{3.2}$$

so that $\pi(B)\,\mathsf{Q}(A,\, A^c \setminus B) \geq \frac{1}{4}\,\mathsf{Q}(A, A^c)^2 \geq \frac{1}{4}\,\Phi(x)^2\,x^2$ when $x = \pi(A)$. Therefore, the $\Phi(x)^2$ in Theorem 3.1 (Average Conductance) can be interpreted as a lower bound on

$\pi(B)\,\mathsf{Q}(A,\,A^c\setminus B)/x^2$. The blocking conductance will replace $\Phi(x)^2$ with the latter quantity and allow the set B to be of size maximizing this product, rather than restricting it to size $\frac{1}{2}\,\mathsf{Q}(A,A^c)$.

Definition 3.1. A function $\phi : [x_0, 1/2] \to [0,1]$ is called an *exterior blocking conductance function* (EBCF) if for all $S \subset \Omega$, $x = \pi(S) \in [x_0,\, 1/2]$ then

$$\forall y \in \left[x,\, \frac{3}{2}x\right] : \Psi_{ext}(S) \geq \phi(\min\{y,\, 1-y\})$$

where

$$\Psi_{ext}(S) = \sup_{\lambda \leq \pi(S)} \min_{\substack{B \subset S^c \\ \pi(B) \leq \lambda}} \frac{\lambda\,\mathsf{Q}(S, S^c \setminus B)}{[\pi(S)\,\pi(S^c)]^2} \;.$$

Note : If $\Psi_{ext}(S) \geq \hat{\phi}(\pi(S))$ some $\hat{\phi}(\cdot)$ and all S, and there exists C such that $\forall x \in [x_0, 1/2]$, $\forall y \in [x,\, 3x/2] : \hat{\phi}(\min\{y, 1-y\}) \leq C\,\hat{\phi}(x)$, then $\phi(x) = \hat{\phi}(x)/C$ is a blocking conductance function.

Because of discreteness of $\pi(S)$, the constraint that $\phi(y) \leq \Psi_{ext}(A)$ is required to extend $\phi(\cdot)$ to the continuous interval $[0, 1/2]$.

The argument before the definition shows that $\phi(x) = \frac{1}{4}\,\Phi(x)^2$ is an example of a blocking conductance function.

A key property of the blocking conductance is that it is a local property, with the value of $\phi(x)$ determined only by sets whose size is of order x. This contrasts to the conductance function $\Phi(x)$, which was determined by all sets of size less than or equal to x. The added generality can be important when the state space has bottlenecks of various sizes, for example bad flow from a single point will not immediately cause poor mixing time bounds.

To further generalize blocking conductance, observe that given a set $A \subset \Omega$ the calculations of (3.2) hold whether $B \subset A$, $B \subset A^c$ or a combination of both. By considering only $B \subset A$ and $B \subset A^c$ we lose at most a factor of two, because $\max\{\pi(B \cap A),\, \pi(B \cap A^c)\} \geq \frac{1}{2}\,\pi(B)$. The most general form of blocking conductance is then

Definition 3.2. A function $\phi(x) : [x_0, 1/2] \to [0,1]$ is called a *blocking conductance function (BCF)* if for all $S \subset \Omega$, $x = \pi(S) \in [x_0,\, 1/2]$ one (but not necessarily both) of the following holds

1. $\forall y \in \left[\frac{1}{2}\,x, x\right] : \Psi_{int}(S) \geq \phi(\max\{x_0,\, y\})$,

2. $\forall y \in \left[x,\, \frac{3}{2}\,x\right] : \Psi_{ext}(S) \geq \phi(\min\{y, 1-y\})$,

where

$$\Psi_{int}(S) = \sup_{\lambda \leq \pi(S)} \min_{\substack{B \subset S \\ \pi(B) \leq \lambda}} \frac{\lambda\, \mathsf{Q}(S \setminus B, S^c)}{[\pi(S)\, \pi(S^c)]^2}$$

and $\Psi_{ext}(S)$ is as in the previous definition.

Again, it suffices to consider a $\hat{\phi}(\cdot)$, find the appropriate C value and set $\phi(x) = \hat{\phi}(\cdot)/C$.

It is reasonable to wonder why we restrict to the case $x \leq 1/2$. However, notice that $\Psi_{int}(S) = \Psi_{ext}(S^c)$ if the definitions are slightly generalized to consider $\min\{x, 1-x\}$ rather than x, so there is no advantage to allowing $x > 1/2$. This is similar to the symmetric conductance function, where $\Phi(A) = \Phi(A^c)$ so there is no reason to define $\Phi(x)$ for $x > 1/2$.

The quantities $\Psi_{int}(S)$ and $\Psi_{ext}(S)$ are measures of vertex and edge bottlenecks. The ratio $\lambda/\pi(S)$ in $\Psi_{int}(S)$ measures the fraction of the interior vertices that can be removed and still have a $\mathsf{Q}(S \setminus B, S^c)/\pi(S)$ fraction of the maximum possible flow $\pi(S)$; likewise for $\Psi_{ext}(S)$. The constraint that $\Psi_{ext}(A) \geq \phi(y)$ is again required to extend $\phi(\cdot)$ from the discrete set of values $\pi(S)$ to the complete interval [0, 1/2].

Then we have

Theorem 3.2. *If $\mathcal{M}$ is a Markov chain (discrete or continuous), $x_0 \leq 1/4$, and $\phi(\cdot)$ is a blocking conductance function, then*

$$\|\rho^t - \pi\|_{TV} \leq \max\left\{\frac{1}{4}\,,\ h(x_0) + \frac{70}{t}\left(\frac{1}{\phi(x_0)} + \int_{x=x_0}^{1/2} \frac{dx}{x\,\phi(x)}\right)\right\}.$$

In particular, if $x_0 = \pi_0$ then $h(x)$ is not required.

Corollary 3.1. *If $\mathcal{M}$ is a Markov chain with $\pi_0 \leq 1/4$, and if $\phi(\cdot)$ is a BCF, then*

$$\tau \leq 140\, K \int_{x=\pi_0}^{1/2} \frac{dx}{x\,\phi(x)}$$

where $K = 1376$ arises from converting between different measures of mixing time.

For any B, $Q(S, B) \leq \pi(B)$ always holds, so if $B \subset S^c$ and $\pi(B) \leq \frac{1}{2}\,\mathsf{Q}(S, S^c)$ then B blocks at most half the flow from S to S^c. Therefore $\Psi(S) \geq \frac{1}{4}\,\Phi(S)^2$, and up to a constant factor, then, Corollary 3.1 is at least as strong as the Average Conductance theorem. The corollary gives an improvement if we can make $\pi(B)$ much larger than $\mathsf{Q}(S, S^c)$ and still maintain high flow.

For example, consider the random walk on the binary n-cube where points are $\mathbf{p} = (p_1, p_2, \ldots, p_n) \in \{0,1\}^n$ and transitions are made by uniformly choosing a coordinate in $[1..n]$ and new value in $\{0,1\}$. Then $\Phi(1/2)$ is minimized by the set $A = \{x \in 2^n : x_n = 0\}$, but $\Psi(A)$ occurs at $\lambda = \frac{1}{4} \gg \frac{1}{4\,n} = \frac{1}{2}\,\mathsf{Q}(A, A^c)$, so we might hope that Corollary 3.1 improves bounds by $O^*(n)$ (which we will later show is true).

Remark : Many variations of Theorem 3.2 are possible by use of different BCF's.

- The condition that $\forall y \in [\pi(A)/2,\, \pi(A)] : \Psi_{int}(A) \geq \phi(y)$ can be relaxed slightly if the condition $\lambda \leq \pi(A)$ in the definition of $\Psi_{int}(A)$ is bounded more strongly. In particular, let $\ell : [x_0, 1/2] \to [0, 1/2]$ satisfy $\forall x : \ell(x) \leq x$, and restrict λ to $\lambda \leq \ell(\pi(A))$. Our previous definition of $\Psi_{int}(A)$ had $\ell(x) = x$. Then it suffices that

$$\forall y \in \left[\pi(A) - \frac{1}{2}\,\ell(\pi(A)),\, \pi(A)\right] : \Psi_{int}(A) \geq \phi(y)\ .$$

  A similar result holds for $\Psi_{ext}(A)$. This generalization will be needed for the proof of Theorem 3.1.

- An *interior BCF* is defined like an *exterior BCF*, but with sets $B \subset A$. Then the theorem and corollary hold as stated but with an interior BCF. An interior BCF can be easier to work with than an exterior BCF, because the condition $\pi(B) \leq \pi(A)$ is satisfied for every $B \subset A$.

- As defined earlier, an *exterior BCF* only uses sets $B \subset A^c$. When considering only an exterior BCF then the $1/\phi(x_0)$ term is not required. Also, the condition $\Psi_{ext}(S) \geq \phi(\min\{y,\, 1-y\})$ can be simplified to $\Psi_{ext}(S) \geq \phi(\min\{y, 1/2\})$, if instead we add an extra term (see proof, this is to compensate for the B_i with $1/2 \in B_i$), to get

$$\|\rho^t - \pi\|_{TV} \leq \max\left\{\frac{1}{4}\,,\ h(x_0) + \frac{70}{t}\left(\int_{x=x_0}^{1/2} \frac{dx}{x\,\phi(x)} + \sup_{x\in[1/3,1/2]} \frac{\ell(x)}{\phi(x)}\right)\right\}\ ,$$

  where $\ell(x)$ is defined at the beginning of these remarks (it suffices to let $\ell(x) = x$).

An exterior BCF has an advantage over interior BCFs because step functions with steps starting at the values given by $\pi(S)$ can be used, for example the conductance function $\Phi(x)$ or the quantity $\Psi_{ext}(S)$.

- A variation on blocking conductance allows $\ell(x) \le 1 - x$ (not just $\ell(x) \le x$). Let

$$\Psi'_{ext}(S) = \sup_{\lambda \le \ell(x)} \min_{\substack{B \subset S^c \\ \pi(B) \le \lambda}} \lambda \, \mathsf{Q}(S, S^c \setminus B) \ .$$

When $\ell(x) = x$ then the restriction of $\lambda \le \ell(x)$ is the same as $\lambda \le x$ before, however if $\ell(x) = 1 - x$ then Ψ' permits much larger sets B to be considered. The disadvantage is that $\phi'(x)$ must satisfy $\forall y \in [\pi(S),\, \pi(S) + \ell(\pi(S))] : \Psi'_{ext}(S) \ge \phi'(y)$. This definition is also better suited for the proof and we are able to improve the constant to

$$\|\rho^t - \pi\|_{TV} \le \max\left\{\frac{1}{4}\,,\ h(x_0) + \frac{28}{t}\left(\frac{1}{\phi(x_0)} + \int_{x=x_0}^{1/2} \frac{x\,dx}{\phi'(x)}\right)\right\} .$$

3.2.2 Proofs

Before giving the proof of the theorem we need some notation. These preliminaries are the same as in [KLM02].

For $u \in \Omega$, let

$$g^t(u) = \frac{\rho^t(u)}{\pi(u)} = \frac{1}{t}\,\frac{\mathsf{p}^{(0)}(u) + \mathsf{p}^{(1)}(u) + \ldots \mathsf{p}^{(t-1)}(u)}{\pi(u)} \ . \tag{3.3}$$

For any 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 $1, 2, \ldots t$. Thus, by linearity of expectations, we get,

$$F(S,T) = \sum_{u \in S} t\, g^t(u)\, \pi(u)\, \mathsf{P}(u,T).$$

Define a probability measure $G(\cdot)$ on Ω by

$$G(A) = \sum_{u \in A} g(u)\, \pi(u).$$

Then,

$$||G - \pi||_{\mathrm{TV}} = \sum_{u \in U} (g(u) - 1)\ \pi(u), \qquad \textit{where } U = \{u : g(u) \ge 1\}. \tag{3.4}$$

The following inequality, which is a modification of a result of [Che00] is crucial to our proof :

Lemma 3.1. *Suppose A, B, C is a partition of Ω into three sets so that there exist positive reals $\alpha > \beta$ and*

$$A \subseteq \{u : g(u) \geq \alpha\}; \quad B \subseteq \{u : g(u) \in [\beta \ , \ \alpha]\}; \quad C \subseteq \{u : g(u) \leq \beta\}.$$

Then, we have

$$\alpha - \beta \leq \frac{1}{t\,\mathsf{Q}(C, A)}.$$

Proof. $F(A, A) + F(A, B \cup C) = F(A, \Omega)$ is the expected number of times we take a step from a state in A (to anywhere) during steps $1, 2, \ldots t$. This is clearly equal to $\mathsf{p}^{(0)}(A) + \mathsf{p}^{(1)}(A) + \ldots + \mathsf{p}^{(t-1)}(A)$. Similarly, $F(A, A) + F(B \cup C, A) = F(\Omega, A)$ is the expected number of times we take a step to a state in A during steps $1, 2, \ldots t$ and this is clearly equal to $\mathsf{p}^{(1)}(A) + \mathsf{p}^{(2)}(A) + \ldots \mathsf{p}^{(t)}(A)$. Subtracting, we get

$$F(B \cup C, A) - F(A, B \cup C) = \mathsf{p}^{(t)}(A) - \mathsf{p}^{(0)}(A) \geq -1. \tag{3.5}$$

From the above, we get

$$\begin{aligned}
-1 &\leq F(B, A) + F(C, A) - F(A, B \cup C) \\
&= \sum_{u \in B} g(u)\, t\pi(u)\, \mathsf{P}(u, A) + \sum_{u \in C} g(u)\, t\pi(u)\, \mathsf{P}(u, A) - \sum_{u \in A} g(u)\, t\pi(u)\, \mathsf{P}(u, B \cup C) \\
&\leq t\alpha \sum_{u \in B} \pi(u)\, \mathsf{P}(u, A) + t\beta \sum_{u \in C} \pi(u)\, \mathsf{P}(u, A) - t\alpha \sum_{u \in A} \pi(u)\, \mathsf{P}(u, B \cup C) \\
&= t\alpha \mathsf{Q}(B, A) + t\beta \mathsf{Q}(C, A) - t\alpha \mathsf{Q}(A, B \cup C) \\
&= t\mathsf{Q}(C, A)(\beta - \alpha),
\end{aligned}$$

where the last line uses time-reversibility. This proves the lemma. □

Proof of Theorem 3.2. Fix the number of iterations t, and write $g(u)$ to indicate $g^t(u)$. Assume there are N states $1, 2, \ldots, N$, and the states are ordered in decreasing order of $g(u)$, that is $g(i) \geq g(j)$ if $i \leq j$; any rule can be used to break ties. We use the notation $[1..i]$ (or $[i..N]$) to indicate $\{1, \ldots, i\}$ (or $\{i, \ldots, N\}$). Also let $a_i = \pi([1..i])$, $i_{min} = \min\{i : a_i \geq x_0\}$, $i_{max} = \max\{i : g(i) > 1\}$ and $\mathcal{I} = \{i_{min}, \ldots, i_{max}\}$.

If $i_{max} < i_{min}$ then $\pi([1..i_{max}]) < x_0$ and so

$$\|\rho^t - \pi\|_{TV} = \pi([1..i_{max}]) - \rho^t([1..i_{max}]) \leq \pi([1..i_{max}]) \leq x_0, \tag{3.6}$$

and the theorem follows because $x_0 \leq 1/4$. Assume $i_{max} \geq i_{min}$. A similar procedure follows if $\pi([i_{max}+1..N]) < 1/4$. Therefore, we may assume that $i_{max} \geq i_{min}$ and that $\pi([1..i_{max}]) \leq 3/4$.

We have

$$\begin{aligned}
\|\rho^t - \pi\|_{TV} &= \sum_{\substack{u\in\Omega \\ g(u)>1}} (g(u)-1)\,\pi(u) \\
&\leq G([1..i_{min}-1]) - \pi([1..i_{min}-1]) + \sum_{i\in\mathcal{I}} \pi([1..i])\,(g(i)-g(i+1))\ .
\end{aligned}$$

Observe that if $j < k$ then

$$\begin{aligned}
\sum_{i=j}^{k} a_i\,(g(i)-g(i+1)) &\leq a_k\,(g(j)-g(k+1)) \\
&\leq a_k\,\frac{1}{t\,\mathsf{Q}([1..j],[k+1..N])}\ ,
\end{aligned}$$

where we used Lemma 3.1 for the final inequality. Therefore, to simplify the variation distance we will group together vertices according to $\phi(\cdot)$, so that the flow $\mathsf{Q}([1..j],[k{+}1..N])$ stays large.

For each $i \in \mathcal{I}$, by definition of Ψ_{int} or Ψ_{ext} applied to $S=[1..i]$ (or $S=[i+1..N]$ if $a_i > 1/2$), let λ_i be the value of λ where the *sup* occurs. Then points can be removed from $[1..i]$ or $[i{+}1..N]$, while still keeping the flow large; in the former case let $\mathcal{B}_i = [a_i - \frac{1}{2}\lambda_i,\, a_i]$ and in the latter let $\mathcal{B}_i = [a_i,\, a_i + \frac{1}{2}\lambda_i]$, in both cases let $B_i = \{j : a_j \in \mathcal{B}_i\}$ and observe that $i \in B_i$.

Fix attention on one i for now. Let the minimal and maximal elements in B_i be $m = \min_{j\in B_i} j$ and $M = \max_{j\in B_i} j$. Extend $\phi(\cdot)$ by letting $\phi(x) = \phi(x_0)$ if $x < x_0$ and $\phi(x) = \phi(1-x)$ for $x > 1/2$, and also let $\Psi(S) = \max\{\Psi_{int}(S), \Psi_{ext}(S)\}$. Suppose that $M = i$, that is λ_i was determined by $\Psi_{int}([1..i])$. Then we have

$$\begin{aligned}
\sum_{j\in B_i} a_j\,(g(j)-g(j+1)) &\leq a_M\,\frac{1}{t\,\mathsf{Q}([1..m],[M+1..N])} \\
&\leq \frac{a_M}{t}\,\frac{\lambda_i}{[a_i(1-a_i)]^2\,\Psi([1..i])} \\
&\leq \frac{a_M}{t}\,\frac{2\,\pi(\mathcal{B}_i)}{[a_i\,(1-a_i)]^2}\,\frac{1}{\phi(a_i)} \\
&\leq \frac{15}{4\,t}\int_{\mathcal{B}_i} \frac{dy}{y\,(1-y)^2\,\phi(y)}\ ,
\end{aligned}$$

where the final expression comes from $a_M = a_i$ and where the factor of $\frac{15}{8}$ is because $a_i\,(1-a_i)^2 \geq \frac{8}{15}\,y(1-y)^2$ for $a_i \in [0,\,3/4]$ and $y \in [a_i - \frac{1}{2}\,\min\{a_i, 1-a_i\},\, a_i]$. The case of $m = i$ (λ_i determined by $\Psi_{ext}([1..i])$) is similar.

Now, let $\mathcal{I}' \subseteq \mathcal{I}$ be such that $\{B_i\}_{i\in\mathcal{I}'}$ forms a minimal cover of the vertices $i \in \mathcal{I}$, i.e. $\mathcal{I} \subseteq \bigcup_{i\in\mathcal{I}'} B_i$. Observe that each point $y \in (0,1)$ is contained in at most two of the $\mathcal{B}_{j\in\mathcal{I}'}$; this is because if three intervals share a point then one interval must be contained in the union of the other two, contradicting minimality of $\mathcal{I}'$.

From these observations it follows that

$$
\begin{aligned}
\|\rho^t - \pi\|_{TV} &\leq G([1..i_{min}-1]) - \pi([1..i_{min}-1]) \\
&\quad + \sum_{i\in\mathcal{I}'} \frac{15}{4\,t} \int_{y\in\mathcal{B}_i} \frac{dy}{y\,\max\{1-y, \frac{1}{4}\}^2\,\phi(\min\{y, 1-y\})} \\
&\leq G([1..i_{min}-1]) - \pi([1..i_{min}-1]) \\
&\quad + \frac{15}{2\,t}\left(\int_{x=x_0/2}^{1/2} \frac{dx}{x(1-x)^2\,\phi(x)} + \int_{x=1/2}^{7/8} \frac{dx}{x\,\max\{1-x, \frac{1}{4}\}^2\,\phi(1-x)}\right) \\
&\leq G([1..i_{min}-1]) - \pi([1..i_{min}-1]) \\
&\quad + \frac{15}{2\,t}\left(4\int_{x=x_0/2}^{1/2} \frac{dx}{x\,\phi(x)} + \frac{16}{3}\int_{x=1/8}^{1/2} \frac{dx}{x\,\phi(x)}\right) \\
&\leq G([1..i_{min}-1]) - \pi([1..i_{min}-1]) + \frac{70}{t}\left(\frac{1}{\phi(x_0)} + \int_{x_0}^{1/2} \frac{dx}{x\,\phi(x)}\right)\,,
\end{aligned}
$$

where the first factor of 2 comes from the number of overlaps by the cover; the upper limit of 7/8 is from the fact that $a_{i_{max}} \leq 3/4$ so the final interval ends before 7/8; and 16/3 is by minimizing $x\max\{1-x, \frac{1}{4}\}^2/(1-x)$ for $x \in (1/2, 7/8)$. □

Proof of Corollary 3.1. Observe that $h(\pi_0) = 0$ so the $h(\cdot)$ term drops out. To eliminate the $\phi(x_0)^{-1}$ term we will show that it can be assumed none of $\mathcal{B}_i$ in the cover extend below π_0. Let $S = [1]$ be the point with maximal $g(u)$. If $\pi(S) \in [\pi_0,\, 2\pi_0)$ then we can assume the BCF uses $\Psi_{ext}([1])$ rather than $\Psi_{int}([1])$, with at most the loss of a factor of 2 because the $\lambda_S^{ext} \geq \pi_0 > \pi(S)/2$. If $\pi(S) \geq 2\pi_0$ then there is no problem because the $\mathcal{B}_i$ in the proof of the theorem have size $\leq \pi(S)/2$, so the $\mathcal{B}_i$ covering $S = [1]$ cannot extend below π_0. □

Proof of Average Conductance Theorem 3.1. We use Corollary 3.1 with external blocking conductance, and $\ell(x)$ as defined in Remark 3.2.1. For $A \subset \Omega$ with $\pi(A) = x$ let $\ell(x) =$

$\lambda = \frac{1}{2}\, x\, \Phi(x) \le x/2$. Now, a set of size λ blocks at most half the flow $\mathsf{Q}(A, A^c)$, so $\mathsf{Q}(A, A^c \setminus B) \ge \frac{1}{2}\, \mathsf{Q}(A, A^c) \ge \frac{1}{2}\, x\, \Phi(x)$. Then $\phi(x) = \frac{1}{4}\, \Phi^2(x)$ is an external BCF. Because $\phi(x)$ is monotonically decreasing then it automatically follows that $\phi(y) \le \Psi_{ext}(A)$.

By Remark 3.2.1 on external BCF's, it remains only to add a term

$$\sup_{x\in[1/3,1/2]} \frac{\ell(x)}{\phi(x)} = \sup_{x\in[1/3,1/2]} \frac{\frac{1}{2}\, x\, \Phi(x)}{\frac{1}{4}\, \Phi^2(x)} \le \frac{1}{\Phi}$$

The theorem then follows (up to constant factors). □

Chapter 4

Geometric Markov Chains

The first case in which we use Average Conductance to show faster mixing are Markov chains whose underlying graphs $G = (E, V)$ have a natural geometric structure.

Roughly speaking, we consider Markov chains where the underlying graph G can be modeled as a convex set K, where if $A \subset V$ then $\pi(A)$ will be proportional to the volume of A and $Q(A, A^c)$ will be proportional to the surface area of the boundary $\partial A \setminus \partial K$. Then the key to bounding the conductance (function) will be to bound the ratio of surface area to volume of cuts; the main tool for doing this will be a type of isoperimetric inequality developed in [KK91], strengthened in [DF91, LS93] and to be further strengthened in this thesis.

4.1 Isoperimetric Inequalities

The key to bounding the conductance function is an *isoperimetric inequality* relating the surface area of a cut to the volume it encloses. First a few definitions.

- A *convex body* is a bounded compact set $K \subset \mathbb{R}^n$ where $\forall x, y \in K, t \in [0, 1]$: $(1-t)x + ty \in K$.
- A function $f : K \to \mathbb{R}$ is *concave* if $\forall x, y \in K, t \in [0, 1] : f[(1-t)x + ty] \geq (1-t)f(x) + tf(y)$.
- A function $F : K \to \mathbb{R}^+$ is *log-concave* if $\log F$ is a concave function on $\text{int}K$, i.e.

$\forall x, y \in K, t \in [0,1] : \log f[(1-t)x + ty] \geq (1-t)\log f(x) + t \log f(y)$. In particular, positive concave functions are log-concave.

- Given a norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^+$, the *dual norm* $\|\cdot\|^*$ is given by

$$\|x\|^* = \sup_{a \in \mathbb{R}^n} \{a \cdot x : \|a\| = 1\} .$$

  For example, when $\|\cdot\| = \|\cdot\|_p$ is the p-norm $\|x\|_p = \left(\sum_{i=1}^n x_i^p\right)^{1/p}$, and q is such that $1/p + 1/q = 1$, then $\|\cdot\|_p^* = \|\cdot\|_q$.

Previous uses of isoperimetry to bound conductance used various refinements of the following theorem of [DF91].

Theorem 4.1. *Let $K \subseteq \mathbb{R}^n$ be a convex body and F a log-concave function on int K. Let $S \subseteq K$ be such that $\partial S \setminus \partial K$ is a piecewise smooth surface σ, with $u(x)$ the Euclidean unit normal to σ at $x \in \sigma$. If $\mu'(S) = \int_\sigma F(x)\|u(x)\|^* \, dx$, $\mu(S) = \int_S F(x)\, dx$ and $\mu(S) \leq \frac{1}{2}\mu(K)$, then*

$$\frac{\mu(S)}{\mu'(S)} \leq \frac{1}{2}\, diam\, K$$

where the diameter diam K is measured with respect to $\|\cdot\|$.

The quantities in the theorem can be better interpreted visually, as in Figure 4.1.

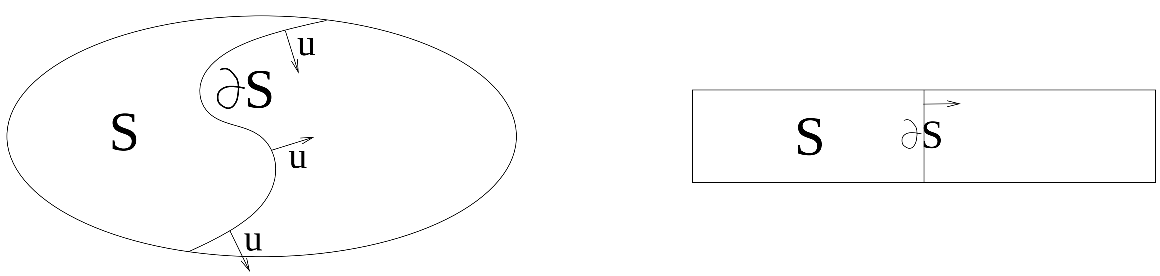

Figure 4.1: Cut surfaces for geometric isoperimetry

When the dual $\|u\|^* \leq c$ along $\partial S \setminus \partial K$ then this theorem can be interpreted as saying the ratio of volume $\mu(S)$ to surface area $\mu'(S)$ is at most $\frac{c}{2}\, diam\, K$, i.e. the surface area is not too small.

Corollary 4.1. *When $F = 1$, $\partial S \setminus \partial K$ is piecewise smooth, and $\|u(z)\|^* \le u_{max}$ then*

$$\frac{\mu_n(S)}{\mu_{n-1}(\partial S \setminus \partial K)} \le \frac{u_{max}}{2}\, diam\, K$$

where μ_n and μ_{n-1} are the n and $n-1$ dimensional Lebesgue measures in $\mathbb{R}^n$.

Taking the norm to be $\|\cdot\|_\infty$ with dual $\|\cdot\|_1$, the space K to be a long thin cylinder like that in Figure 4.1, and S to be half the space as in the figure, then the theorem is tight.

In the problems we consider, the Markov chain $\mathcal{M}$ will have underlying graph (G, V) and subsets $A \subset V$ will map to subsets $S \subset K$; then $Q(A, A^c)$ will be proportional to the surface area $\mu_{n-1}(\partial S \setminus \partial K)$ and $\pi(A)$ will be proportional to the volume $\mu_n(S)$. Then Theorem 4.1 will give a method of bounding Φ.

To bound $\Phi(x)$ a theorem stronger than Theorem 4.1 will be needed. In particular, it will be necessary to bound $Q(A, A^c)$ conditioned on $\pi(A)$ (since $\pi(A) \le x$ in $\Phi(x)$); in the convex body K this will translate into the problem of bounding the surface area $\mu_{n-1}(\partial S \setminus \partial K)$ conditioned on $\mu_n(S)$. Lovász and Kannan [LK99] gave a result of this type for a continuous random walk on convex sets. We will prove a different bound which is appropriate to our applications; one nice feature of our theorem is that it is tight at every value of x, in a sense to be discussed later.

Theorem 4.2. *Under the same conditions as Theorem 4.1, with the added condition that $\mu(S) = x\mu(K) \le \frac{1}{2}\,\mu(K)$, then*

$$\frac{\mu(S)}{\mu'(S)} \le \frac{diam\, K}{H(x)/x}\ . \tag{4.1}$$

where

$$H(x) = \frac{\gamma^2 e^\gamma}{(e^\gamma - 1)^2} \tag{4.2}$$

and $\gamma > 0$ is the unique solution to

$$x = \frac{e^\gamma(\gamma - 1) + 1}{(e^\gamma - 1)^2}\ .$$

Moreover, $H(x)$ is optimal for all $x \in (0, 1/2]$ in that for every x there is an example where the inequality is an equality.

Corollary 4.2. *Under the same conditions as Corollary 4.1, but with* $\mu_n(S) \le x\,\mu_n(K) \le \frac{1}{2}\,\mu_n(K)$, *then*

$$\frac{\mu_n(S)}{\mu_{n-1}(\partial S \setminus \partial K)} \le \frac{u_{max}}{H(x)/x}\, diam\, K \ .$$

where $H(x)$ *is as in Theorem 4.2. In particular,* $H(x) \ge x(2 + \log(1/2x))$.

This extends Theorem 4.1 and also shows that Theorem 4.1 is tight only when $x = 1/2$. Since Theorem 4.2 is optimal then all bounds of the form (4.1) will follow as corollaries to our theorem. The quantity $H(x)$ is difficult to lower bound because of the dependence on γ, however the graph of $H(x)$ demonstrates that

$$\frac{\mu(S)\,\mu(K\setminus S)}{\mu'(S)\,\mu(K)} \le \frac{diam\, K}{4 + \log\left(\frac{1}{4}\,\frac{\mu(K)^2}{\mu(S)\,\mu(K\setminus S)}\right)}\ ,$$

which is a stronger form of a result proven in [KLM02]. The best (graph assisted) approximation we have obtained to (4.2) is $H(x) \ge \sqrt{2\pi}\, I_{gauss}(x)$, or equivalently

$$\frac{\mu(S)}{\mu'(S)} \le \frac{diam\, K}{\sqrt{2\pi}\, I_{gauss}(x)/x}\ ,$$

where $x = \mu(S)/\mu(K) \le 1/2$ and $I_{gauss}(x)$ is the Gaussian isoperimetric function (see Example 4.3).

Bobkov [Bob00] used a Prékopa-Leindler inequality to obtain a related result when the norm is ℓ_2.

Theorem 4.3. *Let* μ *be a log-concave probability measure in* $\mathbb{R}^n$. *For all measurable sets* $S \in \mathbb{R}^n$, *for every point* $x_0 \in \mathbb{R}^n$, *for every number* $r > 0$, *and for* ℓ_2 *norm,*

$$2r\,\mu'(S) \ge \ \mu(S)\,\log\frac{1}{\mu(S)} + (1-\mu(S))\,\log\frac{1}{1-\mu(S)} + \log\mu\{|x - x_0| \le r\}\ .$$

This is not strong enough for our applications because we often require ℓ_∞ norms. When dealing with ℓ_2 norm then Theorem 4.2 and 4.3 are not directly comparable because the latter considers shape as well as size of cuts. For example, taking $r = diam\, K$ and $x_0 \in K$ then Theorem 4.2 is at least twice as strong, but when the object is not very round (e.g. a simplex) then Bobkov's result can be substantially better than Theorem 4.2.

4.1.1 Needles and Isoperimetric Inequalities

The main tool in our proof will be a technique developed by Lovász and Simonovitz [LS90, LS93, KLS99] to reduce n-dimensional isoperimetry problems into 1-dimensional isoperimetry problems. The lemma below is a variation on results in [LS93, KLS99].

A *lower semi-continuous* function is one which is a limit of a monotone increasing sequence of continuous functions. For example, the indicator of an open set, or the negative of the indicator of a closed set.

Lemma 4.1 (Localization Lemma). *Let g and h be lower semi-continuous Lebesgue integrable functions on $\mathbb{R}^n$ such that*

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

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

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

Sketch of Proof. The proof is an application of the Ham Sandwich Theorem. Let $K = \mathbb{R}^n$. Roughly speaking, take an $(n-2)$-dimensional affine hyperplane H (e.g. $x_1 = a_1,\, x_2 = a_2,\, \ldots,\, x_{n-2} = a_{n-2}$ for fixed $a \in \mathbb{R}^{n-2}$) and rotate around the 2 degrees of freedom until the halfspace U above the hyperplane H has $\int_{K\cap U} h(x)\,dx = 0$. Then either $\int_{K\cap U} g(x)\,dx \geq 0$ or $\int_{K\setminus U} g(x)\,dx \geq 0$. Choose the appropriate halfspace $K' = K \cap U$ or $K' = K \setminus U$; repeat this process, bisecting the space each time until only an infinitesimal needle remains. □

The pair $N = \big([a,b],\, \ell(t)^{n-1}\big)$ is often referred to as a *needle* because $\ell(t)^{n-1}$ on $[a,b]$ can be thought of as an infinitesimally narrow truncated cone. In keeping with [KLS99] we let

$$\int_N f = \int_0^{|b-a|} \ell(a + t\,u)^{n-1}\, f(a+tu)\,dt \ ,$$

where $u = (b-a)/|b-a|$. More generally, a *log-concave needle* is of the form $L = ([a,b],\, F(t))$ for F log-concave, and an *exponential needle* is of the form $E = ([a,b],\, e^{\gamma\, t})$, with $\int_L f$ and $\int_E f$ defined as with $\int_N f$. We also let $\mu_L(A) = \int_A F(x)\,dx$ for Lebesgue measurable sets A, and likewise for $\mu_N(A)$ and $\mu_E(A)$.

Observe that if $F(t)$ is log-concave and $\ell(t)$ is linear, then $\tilde{F}(t) = \ell(t)^{n-1} F((1-t)a + tb)$ is log-concave. Then the Localization Lemma reduces a problem on $g = F\,G$ and $h = F\,H$ for F log-concave, to a problem on a log-concave needle $L = ([a, b], \tilde{F}(t))$ with $g = G$ and $h = H$. Therefore most results using the Localization Lemma deal only with log-concave functions as the original n-dimensional problem on log-concave functions reduces to a similar 1-dimensional problem on log-concave needles.

4.1.2 An Isoperimetric Inequality for Log-Concave Functions

In this section we prove Theorem 4.2 to get an isoperimetric inequality sufficient for applying Average Conductance. The proof will proceed in several steps, as in [LS93]. First, a related problem is reduced to a problem on one dimensional needles, we then solve the problem on one dimensional needles, and finally the result is translated into the form of Theorem 4.2.

Theorem 4.4. *Let $K \subseteq \mathbb{R}^n$ be a convex body and F a log-concave function defined on int$K$. Let $S_1,\ S_2 \subseteq K$ be Lebesgue measurable and let $B = K \setminus (S_1 \cup S_2)$. Also let $t \leq dist(S_1,\ S_2)$ and $d \geq diam\, K$, both relative to $\|\cdot\|$.*

Then, given $G : (4,\ \infty) \to \mathbb{R}$ satisfying

1. *$G(1/x)$ is monotonically decreasing for $x \in (0, 1/4)$,*

2. *$x\,G(1/x)$ is monotonically increasing for $x \in (0, 1/4)$,*

it follows that

$$\frac{d}{t}\mu_F(B) \geq \frac{\mu_F(S_1)\,\mu_F(S_2)}{\mu_F(K)}\, G\left(\frac{\mu_F(K)^2}{\mu_F(S_1)\,\mu_F(S_2)}\right) \tag{4.3}$$

holds for all disjoint $S_1,\ S_2 \subseteq K$ if it holds for all one-dimensional exponential needles E with a single $S_1 - B - S_2$ partition, and with $\mu_E(\cdot)$ in place of $\mu_F(\cdot)$.

Proof. Assume a contradiction, i.e. $\exists K,\ S_1,\ S_2,\ B$ with

$$\frac{d}{t}\mu_F(B) < \frac{\mu_F(S_1)\,\mu_F(S_2)}{\mu_F(K)}\, G\left(\frac{\mu_F(K)^2}{\mu_F(S_1)\,\mu_F(S_2)}\right)\ . \tag{4.4}$$

In order to reduce to the needle-like case, the Localization Lemma will require conditions that reinforce the counterexample by decreasing the left side while increasing the right side. The following two conditions will do the job.

- $\mu(B)/\mu(K)$ decreases when changing to needles.
- $x = \mu(S_1)/\mu(K)$ is constant when changing to needles.

These can be written in the form for the Localization Lemma and reduced to the one dimensional case. To do this, we can assume that S_1 and S_2 are closed by taking the closures $\overline{S_1}$ and $\overline{S_2}$. This does not effect t or d, the left side decreases, and the right side increases (by the second monotonicity condition), so this gives another counterexample. But then $B = K \setminus (S_1 \cup S_2)$ is open relative to K, and B and its closure $\overline{B}$ have the same measure. Similarly, K and $int\, K$ are interchangeable because $\mu(\partial K) = 0$ by compactness of K.

Then let

$$\begin{aligned} g(t) &= F(t)\ \left(A\,\mathbf{1}_{int\,K}(t) - \mathbf{1}_{\overline{B}}(t)\right) \text{ where } A = \mu(B)/\mu(K) \\ \text{and } h(t) &= F(t)\ (x\,\mathbf{1}_{int\,K}(t) - \mathbf{1}_{S_1}(t)) \text{ where } x = \mu(S_1)/\mu(K)\ . \end{aligned}$$

These are lower semi-continuous as they are indicators of open sets or negative indicators of closed sets. Then by the Localization Lemma (Lemma 4.1) there is a one-dimensional log-concave needle with the same conditions. Dividing both sides of (4.4) by $\mu_F(K)$ then we see that the condition on $g(t)$ implies the left side of the counterexample decreases, and the condition on $h(t)$ (x constant) implies that $\mu(S_2)/\mu(K)$ increased, so the right side of the counterexample increased by the second monotonicity condition. Also, the needle has smaller diameter (length) than K and larger separation t, so it is still a counterexample.

Moreover, by linearity all norms are equivalent along $\mathbb{R}^1$ up to a constant factor; these constants cancel out when taking d/t, so we can assume the norm on the needle is standard Euclidean length. Without loss assume the needle is $[0, 1]$.

We will now show that this counterexample implies a counterexample for log-concave needles with a single interval $S_1 - B - S_2$.

In general the needle may have many intervals, so we now reduce the general case to a single $S_1 - B - S_2$ interval. To do this we use a trick from [LS93]. Assume the theorem holds for log-concave needles with a single $S_1 - B - S_2$ interval. Consider a maximal interval $[r, s]$ of B. If $\mu([0, r]) \leq \mu([s, 1])$ then color $[0, r]$ red, otherwise color $[s, 1]$ red. Repeat this process over the maximal intervals of B, proceeding from the intervals closest to 0 to those closest to 1. At some point it switches from red on the left ($[0,r]$) to red on the right ($[s,1]$), leaving out an interval $[u,v] \subseteq S_1$ or S_2. Then either S_1 or $S_2 \subseteq [0,u-t_1] \cup [v+t_2,1]$ – where the maximal intervals of B were $[u-t_1,\, u]$ and $[v,\, v+t_2]$ – assume $S_1 \subseteq [0,u-t_1] \cup [v+t_2,1]$. By the single interval case we get

$$\begin{aligned}
\frac{d}{t}\,\mu([u-t_1,u]) &\geq \frac{\mu([0,u-t_1])\,\mu([u,1])}{\mu(K)}\,G\left(\frac{\mu([0,1])^2}{\mu([0,u-t_1])\,\mu([u,1])}\right) \\
&\geq \frac{\mu([0,u-t_1]\cap S_1)\,\mu([u,1]\cup([0,u-t_1]\setminus S_1))}{\mu(K)} \\
&\quad \times G\left(\frac{\mu([0,1])^2}{\mu([0,u-t_1]\cap S_1)\,\mu([u,1]\cup([0,u-t_1]\setminus S_1))}\right) \\
&\geq \frac{\mu([0,u-t_1]\cap S_1)\,\mu(S_2)}{\mu([0,1])}\,G\left(\frac{\mu([0,1])^2}{\mu(S_1)\,\mu(S_2)}\right)
\end{aligned}$$

where the second inequality used $(x-A)(y+A) \leq xy$ when $x<y$, and the second and third inequalities used the first and second monotonicity conditions. Likewise,

$$\frac{d}{t}\,\mu([v,v+t_2]) \geq \frac{\mu([v+t_2,1]\cap S_1)\,\mu(S_2)}{\mu([0,1])}\,G\left(\frac{\mu([0,1])^2}{\mu(S_1)\,\mu(S_2)}\right) .$$

Adding these expressions together gives

$$\frac{d}{t}\,\mu(B) \geq \frac{d}{t}\,(\mu([u-t_1,u]) + \mu([v,v+t_2])) \geq \mu(S_1)\,\mu(S_2)\,G\left(\frac{\mu([0,1])^2}{\mu(S_1)\,\mu(S_2)}\right) ,$$

as desired. If it were $S_2 \subseteq [0,u-t_1] \cup [v+t_2,1]$ then the same steps would hold with S_2.

Now, suppose there is a 1-d counterexample where B consists of a single segment, i.e. the line is of the form $S_1 - B - S_2$. Assume $S_1 = [0,a]$, $B = (a,b)$, $S_2 = [b,1]$ and let $\log \tilde{F}(t)$ be the line $\log \tilde{F}(t) = A + \gamma\, t$ passing through the points $(a,\, \log F(a))$ and $(b,\, \log F(b))$. By the log-concavity of $F(t)$, it follows that $\tilde{F}(t) \leq F(t)$ in B and $\tilde{F}(t) \geq F(t)$ in $S_1 \cup S_2$. We now show that $\tilde{F}(t)$ will give a counterexample to the exponential needle problem with the same $S_1 - B - S_2$.

Let $\mu_{\tilde{F}}(A) = \int_A \tilde{F}(t)\,dt$ be the measure induced on subsets $A \subseteq [0,1]$ by $\tilde{F}(t)$, and similarly let $\mu_F(A) = \mu(A) = \int_A F(t)\,dt$. Then $\mu_{\tilde{F}}(B) \leq \mu_F(B)$ so the left side of the counterexample decreased. Now, $\mu_{\tilde{F}}(S_1) \geq \mu_F(S_1)$ and $\mu_{\tilde{F}}(S_2) \geq \mu_F(S_2)$, and also $\mu_{\tilde{F}}(S_1)/\mu_{\tilde{F}}(K) \geq \mu_F(S_1)/\mu_F(K)$ or $\mu_{\tilde{F}}(S_2)/\mu_{\tilde{F}}(K) \geq \mu_F(S_2)/\mu_F(K)$. Then $\mu(S_1)\,\mu(S_2)/\mu(K)$ increases in going from F to $\tilde{F}$, so by the second monotonicity condition then the right side of the counterexample increases.

Then there is a one dimensional single interval counterexample with an exponential needle, contradicting the assumption. □

We have reduced to exponential needles on $[0, 1]$ with a single B interval. Assume the needle is $E = ([0, 1], e^{\gamma t})$, and the interval is partitioned into three pieces – $S_1 = [0, s]$, $B = [s, s+t]$, $S_2 = [s+t, 1]$ – where $\mu(S_1) = \int_0^s e^{\gamma y}\,dy$ and likewise for B and S_2.

Theorem 4.5. *Let G be a function $G : (4, \infty) \to \mathbb{R}$ satisfying the two monotonicity conditions of Theorem 4.4. Then if*

$$\forall \gamma > 0 : \; x(1-x)G\left(\frac{1}{x(1-x)}\right) \le \frac{\gamma^2 e^{\gamma}}{(e^{\gamma}-1)^2} \tag{4.5}$$

where

$$x = \frac{e^{\gamma}(\gamma - 1) + 1}{(e^{\gamma}-1)^2} \in (0, 1/2) \ ,$$

then the conditions of Theorem 4.4 are satisfied. Moreover, (4.5) is both necessary and sufficient for this to be true.

Proof. In the following work extend $G : (4, \infty) \to \mathbb{R}^+$ to $G : [4, \infty) \to \mathbb{R}^+$ by letting $G(4) = \lim_{t \to 4^+} G(t)$. This is well defined because G is monotonic. In the work below we will not worry about $\gamma = 0$, but this case can easily be worked out with the same methods.

We first show that the second condition implies the first.

Consider a counterexample with $\gamma > 0$

$$\frac{\int_s^{s+t} e^{\gamma y} dy}{t} < \frac{\left(\int_0^s e^{\gamma y} dy\right)\left(\int_{s+t}^1 e^{\gamma y} dy\right)}{\int_0^1 e^{\gamma y} dy} G\left(\frac{\left(\int_0^1 e^{\gamma y} dy\right)^2}{\left(\int_0^s e^{\gamma y} dy\right)\left(\int_{s+t}^1 e^{\gamma y} dy\right)}\right) \ . \tag{4.6}$$

The right side is decreasing in t by the second monotonicity condition, while the left side is increasing in t because it is the average of the increasing function $e^{\gamma y}$. Then taking $t \to 0^+$ on both sides gives another counterexample :

$$e^{\gamma s} < \mu([0, 1])\, x(1-x)\, G(1/x(1-x))$$

where

$$x = \frac{\mu([0, s])}{\mu([0, 1])} = \frac{\gamma^{-1}(e^{\gamma s} - 1)}{\mu([0, 1])} \ .$$

The case $\gamma < 0$ is similar but with $t^{-1} \int_{s-t}^s e^{\gamma y} dy$ on the left side of (4.6).

Solving for $e^{\gamma s}$ in terms of x and substituting into the counterexample gives

$$\gamma x + \frac{\gamma}{e^\gamma - 1} < x(1-x)\, G(1/x(1-x)) \ . \tag{4.7}$$

Fix x and minimize the left side with respect to γ.

$$\frac{\partial}{\partial \gamma}\left(\gamma x + \frac{\gamma}{e^\gamma - 1}\right) = x + \frac{e^\gamma - 1 - \gamma e^\gamma}{(e^\gamma - 1)^2}$$

This is increasing in γ because $\frac{d}{d\gamma} > 0$ except at $\gamma = 0$, so the optima is an absolute minima, i.e. the minimum occurs at the solution to

$$x = \frac{e^\gamma(\gamma - 1) + 1}{(e^\gamma - 1)^2} \in (0,1) \ . \tag{4.8}$$

Observe that for $\gamma \in (-\infty, \infty)$ then (4.8) is a bijection onto $x \in (0,1)$. Since the solution to (4.8) is the minimum of the left side in (4.7) then there is another counterexample with

$$\gamma \frac{e^\gamma(\gamma - 1) + 1}{(e^\gamma - 1)^2} + \frac{\gamma}{e^\gamma - 1} < x(1-x)\, G\left(\frac{1}{x(1-x)}\right) \ .$$

This simplifies to give a counterexample to (4.5). Notice that $\gamma^2\, e^\gamma/(e^\gamma - 1)^2$ is the same for $\pm\gamma$, and that $x(-\gamma) = 1 - x$, so it suffices to consider $\gamma > 0$.

To show that the first condition implies the second, observe that the above work showed that given x then taking $t \to 0^+$ the γ given by (4.8) satisfies the second condition in the theorem. Then the second condition follows. □

Proof of Theorem 4.2. The first part of the proof is identical to that in [DF91], but with our stronger bound from Theorem 4.5. For completeness we give the proof from [DF91].

By considering the limit of an appropriate sequence of simplicial approximations, it clearly suffices to prove the theorem for σ a "simplicial surface", i.e. one whose "pieces" are $(n-1)$-dimensional simplexes. For small $t > 0$, let B be the closed $\frac{1}{2}\,t$-neighborhood of such a surface σ. Consider a simplicial piece $\sigma' \subseteq \sigma$, with normal u and surface integral $\alpha = \int_{\sigma'} F(x)\, dx$. The measure of B around σ' is then approximately $h\alpha$, where

$$h = \max\{uz : \|z\| = t\} = \|u\|^*\, t \ .$$

Thus the measure of this portion of B is $t\alpha\|u\|^* + o(t)$ and hence, since u is constant on each such σ', $\mu(B) = t\mu'(S) + o(t)$.

Now, $G(1/y)$ from Theorem 4.5 satisfies both monotonicity conditions of Theorem 4.4 so if $S_1 = S \setminus B$, and $S_2 = K \setminus (B \cup S)$, then we have

$$\frac{d}{t}\,\mu(B) \geq \frac{\mu(S_1)\,\mu(S_2)}{\mu(K)}\,G\left(\frac{\mu(K)^2}{\mu(S_1)\,\mu(S_2)}\right)\ .$$

The first part of the theorem follows by letting $t \to 0^+$.

For the second part, the construction in the proof of Theorem 4.5 shows that if $K = [0,1]$ then $F(y) = e^{\gamma\,y}$ is tight at $x = [e^\gamma(\gamma-1)+1]/(e^\gamma-1)^2$ as $t \to 0^+$. This generalizes to n dimensions by taking $K = [0,1]^n$ and $F(y) = e^{\gamma\,y_1}$, where y_1 is the first coordinate of y. $\square$

For brevity we will use the notation

$$H(x) = \frac{\gamma^2 e^\gamma}{(e^\gamma-1)^2} \quad \textit{where} \quad x = \frac{e^\gamma(\gamma-1)+1}{(e^\gamma-1)^2} \in (0,1/2)\ ,$$

to denote the upper bound on $x(1-x)G(1/x(1-x))$.

Remark : The proof of Theorem 4.5 shows that given x, then letting γ be the unique solution to $x = (e^\gamma\,(\gamma-1)+1)/(e^\gamma-1)^2$, $\|\cdot\| = \|\cdot\|_\infty$, $F = e^{\gamma\,y}$, and $K = [0,1]^n$ then the inequality is an equality. When $x = 1/2$ then $F = 1$ and this gives the same example of equality as that given in [DF91].

Remark : The γ in Theorem 4.2 can be interpreted as the slope of $H(x)$. To see this, observe that

$$\begin{aligned}
\frac{d}{dx}H(x) &= \frac{d}{d\gamma}\left[\frac{\gamma^2 e^\gamma}{(e^\gamma-1)^2}\right]\,\left[\frac{dx}{d\gamma}\right]^{-1} \\
&= \frac{\frac{(2\gamma e^\gamma+\gamma^2 e^\gamma)(e^\gamma-1)^2-\gamma^2 e^\gamma 2e^\gamma(e^\gamma-1)}{(e^\gamma-1)^4}}{\frac{(e^\gamma+e^\gamma(\gamma-1))(e^\gamma-1)^2-(e^\gamma(\gamma-1)+1)2e^\gamma(e^\gamma-1)}{(e^\gamma-1)^4}} \\
&= \frac{\gamma e^\gamma(e^\gamma-1)\left[(2+\gamma)(e^\gamma-1)-2\gamma e^\gamma\right]}{e^\gamma(e^\gamma-1)\left[\gamma(e^\gamma-1)-2(e^\gamma(\gamma-1)+1)\right]} \\
&= \gamma\ .
\end{aligned}$$

Example 4.1. Consider the n-dimensional hypercube $[0,1]^n$ with $F = 1$, ℓ_∞ norm and $S \subset K$ required to have flat faces (i.e. $\|u\|_1 = 1$). The earlier remark mentions that this is tight at $x = 1/2$. However, when $x \to 0^+$ then Bollobás and Leader [BL91b] showed that the surface area is minimized by taking small sub-cubes $[0,\epsilon]^n$ in a corner. This has

$$\frac{\mu(S)}{\mu'(S)} = \frac{x}{n\,x^{(n-1)/n}} = \sqrt[n]{x}/n \ ,$$

while Theorem 4.2 gives

$$\frac{\mu(S)}{\mu'(S)} \leq \frac{diam\,K}{H(x)/x} = \frac{e^\gamma(\gamma-1)+1}{\gamma^2\,e^\gamma} \ .$$

Taking $x \to 0^+$ ($\gamma \to \infty$) then the ratio of these two bounds is

$$\frac{\sqrt[n]{\frac{e^\gamma(\gamma-1)+1}{(e^\gamma-1)^2}}}{\frac{e^\gamma(\gamma-1)+1}{\gamma^2\,e^\gamma}} \xrightarrow{\gamma\to\infty} 0 \ .$$

So at least for the hypercube we see that the bound of Theorem 4.2 can be arbitrarily bad for small x. In the next section we will prove a version of Theorem 4.2 for the uniform distribution $F = 1$ and show that it is asymptotically tight for $x \to 0^+$ in the Bollobás and Leader example.

Theorem 4.2 gives an optimal bound, but it seems impossible to write H in closed form. Below we show a few good approximations to $H(x)$, and hence good lower bounds for Theorem 4.2.

Example 4.2. Consider bounds of the form

$$H(x)/x \geq A + \log_\alpha(1/x) \tag{4.9}$$

where $x \in (0, 1/2)$.

Use the notation of (4.5), where $x = \frac{e^\gamma(\gamma-1)+1}{(e^\gamma-1)^2}$. First consider the base α. Dividing both sides of (4.9) by $\ln(1/x)$, then

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

Taking $\gamma \to \infty$, this shows that $1 \geq 1/\ln\alpha$, so $\alpha \geq e$. From now on the log is base e.

Next consider the constant A. Then (4.9) is equivalent to

$$\frac{\gamma^2 e^\gamma}{e^\gamma(\gamma-1)+1} - \ln\left(\frac{(e^\gamma-1)^2}{e^\gamma(\gamma-1)+1}\right) \geq A$$

Taking $\frac{d}{d\gamma}$ shows that the left side of this equation is increasing in γ, so letting $\gamma \to 0^+$ this reduces to $2 - \ln 2 \geq A$.

Now the conditions of Theorem 4.4. The first condition holds for any constant A. For the second condition it suffices that $\frac{d}{dx}H(x) \geq 0$, that is

$$\frac{d}{dx}\left[x\left(A + \ln(1/x)\right)\right] = A + \ln(1/x) - 1 \geq 0 \ .$$

In particular $A \geq 1 - \ln(1/x)$, so any $A \geq 1 - \ln 2$ suffices.

From this we get that the isoperimetric inequality holds for

$$H(x)/x = 2 - \ln 2 + \ln(1/x) \ .$$

Kannan [Kan] previously showed a result similar to Theorem 4.2, but with $H(x)/x = 1 + \ln(1/x)$. This would be our lower bound on A if we required the second monotonicity condition in Theorem 4.4 to hold for $x \in (0, 1)$, rather than the weaker $x \in (0, 1/4)$.

Example 4.3. An extremely good bound can be found with a Gaussian type isoperimetric inequality as in [Bob00]. Let $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$ be the standard normal distribution (this is not the conductance function), $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (note $\frac{dN}{dx} = \varphi(x)$), and $I_{gauss}(x) = \varphi \circ N^{-1}(x)$. Then a Gaussian type inequality has the form

$$\frac{d}{t}\frac{\mu(B)}{\mu(K)} \geq c\, I_{gauss}(x) \tag{4.10}$$

for some constant c.

From Example 4.2 it suffices to show that $cI_{gauss}(x)/x \leq 2 + \log(1/2x)$. However, it is known that $I_{gauss}(x) \leq C\, x\sqrt{\log(1/x)}$ for some constant C, and so a Gaussian type inequality will hold for some c.

Determining the optimal value of c is difficult. At this point we do not have a complete proof of the optimal value. However, in the following we reduce the problem to a (possibly) simpler one; this later problem can be easily graphed on a computer, and via this graph we are able to verify that $c = \sqrt{2\pi}$ is optimal.

From Theorem 4.5, it suffices to show

$$c\, I_{gauss}(x) \leq \frac{\gamma^2}{(e^\gamma - 1)^2} \qquad for \qquad x = \frac{e^\gamma(\gamma - 1) + 1}{(e^\gamma - 1)^2} \ .$$

Consider the endpoints. For $x \to 0^+$ ($\gamma \to \infty$), we have $\frac{\gamma^2}{(e^\gamma-1)^2} \xrightarrow{\gamma\to\infty} 0$ and $I_{gauss}(x) \xrightarrow{x\to 0}$ 0, so the inequality holds for any c. When $x \to 1/2$ ($\gamma \to 0^+$) then $\frac{\gamma^2}{(e^\gamma-1)^2} \xrightarrow{\gamma\to 0^+} 1$ and $I_{gauss}(1/2) = 1/\sqrt{2\pi}$, so the inequality holds for $c \leq \sqrt{2\pi}$. Assume $c = \sqrt{2\pi}$.

To show the inequality for general $x \in (0, 1/2)$ consider $\frac{d}{dx}\left[H(x) - c\, I_{gauss}(x)\right]$.

Then

$$\begin{aligned}
\frac{d}{dx}\varphi\left[N^{-1}(x)\right] &= \varphi'\left[N^{-1}(x)\right]\,\frac{d}{dx}N^{-1}(x) \\
\varphi'(x) &= \frac{1}{\sqrt{2\pi}}(-x)e^{-x^2/2} = -x\varphi(x) \\
\frac{d}{dx}N^{-1}(x) &= 1/N'\left[N^{-1}(x)\right] = 1/\varphi\left[N^{-1}(x)\right] \ .
\end{aligned}$$

So

$$\frac{d}{dx}\varphi\left[N^{-1}(x)\right] = -N^{-1}(x) \ .$$

From (4.9) we know that $\frac{d}{dx}H(x) = \gamma$, so

$$\frac{d}{dx}\left[H(x) - c\, I_{gauss}(x)\right] = \gamma + c\, N^{-1}(x) \ .$$

Therefore, the extreme points occur when $\gamma + c\, N^{-1}(x) = 0$, or equivalently when $N(-\gamma/c) - x = 0$. This latter form can be plotted for $c = \sqrt{2\pi}$ and has a single root when $x \in (0, 1/2)$, at $\gamma \sim 3.5$. Therefore, the original function is either strictly positive or strictly negative, as it has only one extreme point for $x \in (0, 1/2)$ and is 0 at the endpoints. The graph of $N(-\gamma/c) - x$ is negative for $\gamma >\sim 3.5$ which means that $\frac{d}{dx}$ is positive for $\gamma >\sim 3.5$, i.e. x small, and shows that $\frac{d}{dx} > 0$ for small x and hence for all $x \in (0, 1/2)$.

This shows that

$$\frac{d}{t}\frac{\mu(B)}{\mu(K)} \geq \sqrt{2\pi}\, I_{gauss}(x) \ .$$

This is tight at both endpoints and extremely close to optimal at the intermediate points.

4.1.3 Isoperimetric Inequalities for Uniform Distributions

When the distribution F is uniform over K then the results from the previous section can be strengthened slightly.

Theorem 4.6. *The same results as in Theorem 4.4 hold for $\mu(S) = Vol_n(S)$ (i.e. $F = 1$) and reduction to a single interval truncated pyramid (linear) case (i.e. $\ell(t) = (\alpha + \beta t)^{n-1}$).*

Proof. The proof is identical to before without the step of reducing to an exponential needle. □

Theorem 4.7. *Let G be a function $G : (4, \infty) \to \mathbb{R}$, then if*

$$\forall \gamma > 0 : x(1-x)\, G\left(\frac{1}{x(1-x)}\right) \leq \frac{\gamma\, n}{(1+\gamma)^n - 1} \left[\frac{(1+\gamma)^{n-1}\gamma(n-1)}{(1+\gamma)^{n-1}-1}\right]^{1-1/n} \tag{4.11}$$

$$\textit{where} \;\; x = \frac{(1+\gamma)^{n-1}\left[\gamma(n-1)-1\right]+1}{\left[(1+\gamma)^{n-1}-1\right]\left[(1+\gamma)^n-1\right]} \in (0, 1/2),$$

then the conditions of Theorem 4.6 are satisfied. Moreover, (4.11 is both a necessary and sufficient condition for this to be true.

Proof. The proof will follow the same steps as the general case. As before, extend G to $G : [4, \infty) \to \mathbb{R}^+$ and ignore the case $\gamma = 0$.

We first show that the second condition implies the first.

The truncated pyramid is given by $(\alpha + \beta\, y)^{n-1}$; this reduces to two cases, $(1 + \gamma y)^{n-1}$ or y^{n-1}, by dividing by α^{n-1} when $\alpha \neq 0$ or β^{n-1} when $\alpha = 0$. Assume $\gamma \geq 0$, the case $\gamma < 0$ follows similarly.

First deal with the harder case, $(1 + \gamma y)^{n-1}$ and $\gamma \in [-1, \infty)$. Then $\mu(S_1) = \int_0^s (1 + \gamma y)^{n-1} dy$ and likewise for B and S_2. Simplify notation by setting $x = \mu([0, s])/\mu([0, 1])$.

Consider a counterexample with $\gamma > 0$

$$\frac{\int_s^{s+t}(1+\gamma y)^{n-1}dy}{t} < x \int_{s+t}^{1} (1+\gamma y)^{n-1} dy\, G\left(\frac{\int_0^1 (1+\gamma y)^{n-1}dy}{x\, \int_{s+t}^1 (1+\gamma y)^{n-1} dy}\right) . \tag{4.12}$$

The right side is decreasing in t by the second monotonicity condition, while the left side is increasing in t because it is the average of the increasing function $(1 + \gamma\, y)^{n-1}$. Then taking $t \to 0^+$ on both sides gives another counterexample :

$$(1+\gamma s)^{n-1} < \mu([0,1]) x(1-x)\, G(1/x(1-x))$$

where

$$x = \frac{\mu([0,s])}{\mu([0,1])} = \frac{(1+\gamma s)^n - 1}{(1+\gamma)^n - 1} \; .$$

The case $-1 < \gamma < 0$ is similar but with $t^{-1} \int_s^{s+t} (1 + \gamma y)^{n-1} dy$ on the left side of (4.12).

Solving for $(1+\gamma s)^{n-1}$ in terms of x and substituting into the counterexample gives

$$\frac{\gamma n \left\{1 + x\left[(1+\gamma)^n - 1\right]\right\}^{(n-1)/n}}{(1+\gamma)^n - 1} < x(1-x)\, G(1/x(1-x)) \ . \tag{4.13}$$

Fix x and n and minimize the left side with respect to γ. Simplify by setting $u = (1+\gamma)^n - 1$.

$$\begin{aligned} \frac{\partial}{\partial \gamma} \frac{\gamma n(1+xu)^{(n-1)/n}}{u} &= \frac{\left[n(1+x\,u)^{1-1/n} + \gamma n \frac{n-1}{n}(1+x\,u)^{-1/n} x \frac{\partial u}{\partial \gamma}\right] u - \gamma n(1+x\,u)^{1-1/n} \frac{\partial u}{\partial \gamma}}{u^2} \\ &= \frac{n(1+x\,u)^{-1/n}}{u^2} \left\{ \left[u - \gamma \frac{\partial u}{\partial \gamma}\right] + x\left[u^2 - \frac{u\gamma}{n}\frac{\partial u}{\partial \gamma}\right] \right\} \end{aligned}$$

This is increasing in γ, so the optima is an absolute minima, i.e. the minimum occurs at the solution to

$$x = \frac{(1+\gamma)^{n-1}\left[\gamma(n-1) - 1\right] + 1}{\left[(1+\gamma)^{n-1} - 1\right]\left[(1+\gamma)^n - 1\right]} \in (0,1) \ . \tag{4.14}$$

Observe that for $\gamma \in (-1, \infty)$ then (4.14) is a bijection onto $x \in (0,1)$. Since the solution to (4.14) is the minimum of the left side in (4.13), then substituting for x gives another counterexample with

$$\frac{\gamma n \left\{1 + \frac{(1+\gamma)^{n-1}[\gamma(n-1)-1]+1}{[(1+\gamma)^{n-1}-1][(1+\gamma)^n-1]}\left[(1+\gamma)^n - 1\right]\right\}^{(n-1)/n}}{(1+\gamma)^n - 1} < x(1-x)\, G(1/x(1-x)) \ .$$

This simplifies to give a counterexample to the second condition.

A similar proof for distribution y^{n-1} shows $G(x) \leq n\sqrt[n]{x}$, which is a weaker condition than the second condition in the theorem.

To show that the first condition implies the second, observe that the above work showed that given x then taking $t \to 0^+$ the γ given by (4.14) satisfies the second condition in the theorem. Then the second condition follows. □

Note : Two points in this proof were computer assisted. The statement that the quantity above (4.14) is increasing, and the reduction from $\gamma \in (-1, \infty)$ to $\gamma > 0$ were both done by graphing on Mathematica. This makes the proof incomplete, however we do not use Theorem 4.7 outside of Section 4.1.3, so this is not a major problem.

As before, this gives a result on surface area and volume. In the next two examples we let $H(x) = (1-x)\, G(1/x(1-x))$ where x and G are as in the previous theorem.

Example 4.4. It was shown in [BL91b] that on small sets the hypercube with flat faces has isoperimetric inequality $H(x) = n\sqrt[n]{1/x}$ with $\|\cdot\| = \|\cdot\|_\infty$. This was the bound in our theorem when we only considered y^{n-1}, i.e. with cones and not truncated cones. However, in general Theorem 4.7 behaves asymptotically like the hypercube when $x \to 0^+$:

$$\begin{aligned}
\lim_{x\to 0} \frac{n\sqrt[n]{x^{-1}}}{H(x)/x} &= \lim_{\gamma\to\infty} \frac{n\sqrt[n]{\frac{[(1+\gamma)^{n-1}-1][(1+\gamma)^n-1]}{(1+\gamma)^{n-1}[\gamma(n-1)-1]+1}}}{\gamma n\sqrt[n]{(1+\gamma)^{n-1}-1}\frac{[(1+\gamma)^{n-1}\gamma(n-1)]^{1-1/n}}{(1+\gamma)^{n-1}[\gamma(n-1)-1]+1}} \\
&= \lim_{\gamma\to\infty} \frac{\sqrt[n]{(1+\gamma)^n-1}}{\gamma}\left\{\frac{(1+\gamma)^{n-1}[\gamma(n-1)-1]+1}{(1+\gamma)^{n-1}\gamma(n-1)}\right\}^{1-1/n} \\
&= 1
\end{aligned}$$

Example 4.5. Bollobás and Leader [BL91b] conjectured that the subsets with smallest boundary in the unit hypercube are "cylinders" of the form :

$$\exists r \in \{1,\ldots,n\},\ a \in [0,n] : B = \{x \in I^n : \sum_{i=1}^{r} x_i^2 \le a\}$$

where *smallest boundary* is exactly $\mu'(S)$ with norm $\|\cdot\|_2$. In other words, they conjecture that $H(x)$ is determined by the cylinders.

Consider the limit as $x \to 0^+$, so that the smallest term in the conjecture is $r = n$, i.e. a partial sphere embedded in the corner of the hypercube. Then

$$\begin{aligned}
x &= 2^{-n}Vol_n(\delta) \\
&= \left(\frac{\delta}{2}\right)^n \begin{cases} \frac{\pi^{n/2}}{(n/2)!} & if\ n\ is\ even \\ \frac{\pi^{(n-1)/2}\,2^{(n+1)/2}}{1\cdot 3\cdots n} & if\ n\ is\ odd \end{cases} \\
&\approx \left(\frac{\pi\, e}{2\, n}\right)^{n/2} \frac{\delta^n}{\sqrt{\pi n}}
\end{aligned}$$

where the final term used Stirling's Formula $n! \approx (n/e)^n\sqrt{2\pi n}$.

By the generalized Cavalier principle, or a simple bit of calculus, we see

$$\frac{SA_n(\delta)}{Vol_n(\delta)} = \frac{n}{\delta} \approx \sqrt{\frac{\pi\, e}{2\sqrt[n]{\pi n}}}\,\frac{\sqrt{n}}{\sqrt[n]{x}}\ .$$

By Example 4.4 we saw that $\lim_{x\to 0}\frac{H(x)}{n/\sqrt[n]{x}} = 1$. Now, $diam_2\, K = \sqrt{n}$, so $H(x)$ is decreased by a factor of $\sqrt{n}$ from the $\|\cdot\|_\infty$ case in the previous example. Then

$$\lim_{x\to 0^+}\frac{cylinder\ F}{my\ bound} \approx \sqrt{\frac{\pi\, e}{2\sqrt[n]{\pi\, n}}} \approx 2\ .$$

While this doesn't answer the conjecture it shows that at least for small sets the conjecture is within a factor of about 2 of optimal, and even closer when the dimension n is small.

4.2 Edge-Isoperimetry

The connection between Theorem 4.2 and edge-isoperimetry can be seen clearly by an example. This is done by associating vertices of graphs with simplexes, such that two vertices are adjacent exactly when their associated simplexes share a face. The technique is similar to that developed in [DFK91, KK91] to bound the cutset of a graph, our contribution is in extending these inequalities to edge-isoperimetry.

Our toy example throughout this chapter will be a random walk on a grid. A more practical example will be given at the end of the chapter.

Example 4.6. Let $G = (E, V)$ be the grid $[k]^n$, the n-dimensional cube of side length k, and write the vertices V of G in Cartesian product form so that

$$V = \{v = (v_1, v_2, \ldots, v_n) : v_i \in [1, \ldots, k]\}$$

To each vertex $v \in G$ associate the polytope

$$P(v) = \{x \in \mathbb{R}^n : v_i - 1 \leq x_i \leq v_i \text{ for all } i \in [1, \ldots n]\} .$$

and denote the image of G by $\Omega = \bigcup_{v \in G} P(v) = [0, k]^n \subset \mathbb{R}^n$.

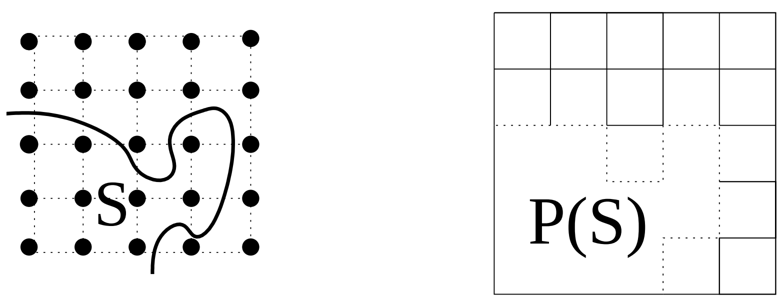

Figure 4.2: Mapping a graph into a convex body

Properties such as adjacency and cut size in G carry over well to Ω. Cuts $S \subseteq V$ with $|S|/|V| \leq x$ map to cuts $P(S)$ of Ω with

$$vol_n P(S)/vol_n \Omega = |S|/|V| \leq x .$$

Two vertices $v^1, v^2 \in G$ are adjacent if and only if $P(v^1)$ and $P(v^2)$ intersect at a face, and

$$|Cut(S)| = vol_{n-1} (\partial P(S) \cap int(\Omega)) \tag{4.15}$$

The right hand side of (4.15) is just $\mu'(P(S))$ when $F = 1$ and the norm is $\|\cdot\|_\infty$ (observe $\|u\|^* = 1$ for $F = 1$), while $\mu(P(S)) = vol_n P(S)$. This suggests the use of Theorem 4.2 to bound $|Cut(S)|$.

$$\frac{\mu(P(S))}{\mu'(P(S))} = \frac{vol_n P(S)}{vol_{n-1}\left(\partial P(S) \cap int(\Omega)\right)} \leq \frac{diam_\infty \Omega}{2 + \log(1/2x)} \ . \tag{4.16}$$

Algebraic manipulation of (4.15) and (4.16), along with $diam\Omega = k$ and $vol_n P(S) = |S|$ give

$$\frac{|Cut(S)|}{|S|} \geq \frac{1}{k}(2 + \log(1/2x)) \text{ when } |S| \leq x|G| \ . \tag{4.17}$$

This is correct when $x = 1/2$, and is always within a factor e of the optimal (in terms of log only) inequality [BL91b]

$$\frac{|Cut(S)|}{|S|} \geq \frac{2}{k} \log_2(1/x) \text{ when } |S| \leq x|G| \ .$$

The properties important for use of the isoperimetric inequality in the previous example are captured by the following definition.

Definition 4.1. We say an undirected weighted graph $G = (E, V)$ with vertex weights $\pi(v)$ and edge weights $\mathsf{Q}(x, y)$ is a *geometric graph* if there is a mapping

$$\phi : V \to \text{Simplexes in } \mathbb{R}^n$$

such that $K = \phi(V)$ is a convex body in $\mathbb{R}^n$, the simplexes are disjoint except possibly along the faces, vertex weights are preserved (i.e. $\forall v \in V : \mu_n(\phi(v)) \propto \pi(v)$), and edge weights are preserved (i.e. $\forall v_1,\, v_2 \in V : \mu_{n-1}(\phi(v_1) \cap \phi(v_2)) \propto \mathsf{Q}(v_1, v_2)$).

With this we can generalize the result of Example 4.6.

Theorem 4.8. *Let $G = (E, V)$ be a geometric (weighted) graph with vertex constant of proportionality A, edge constant B, diameter $diam\, K$ in norm $\|\cdot\|$, and $u(x)$ the Euclidean unit normal to $\partial\phi(v)$ at x satisfying $\|u\|^* \leq u_{max}$ for all x along $\partial\phi(v)$. Then the graph has cutset expansion*

$$\frac{\mathsf{Q}(S, S^c)}{\pi(S)} \geq \gamma \qquad \textit{when } \pi(S) \leq \frac{1}{2}\,\pi(V)$$

and edge-isoperimetric inequality

$$\frac{\mathsf{Q}(S, S^c)}{\pi(S)} \geq \frac{2 + \log(1/2x)}{2}\,\gamma \qquad \textit{when } \pi(S) \leq x\pi(V) \leq \frac{1}{2}\,\pi(V)\ ,$$

where

$$\gamma = \frac{A}{B}\,\frac{2}{u_{max}\,diam\,K}\ .$$

Proof. By definition of a geometric graph, there is a model of the graph as a convex body. Suppose $S \subset V$, then by the construction $\mu_n(\phi(S)) = A\,\pi(S)$ and $\mu_{n-1}(\phi(S) \cap \phi(S^c)) = B\,\mathsf{Q}(S, S^c)$. Corollary 4.1 implies that

$$\begin{aligned} \frac{\mathsf{Q}(S,S^c)}{\pi(S)} &= \frac{A}{B}\,\frac{\mu_{n-1}(\partial\phi(S) \setminus \partial K)}{\mu_n(\phi(S))} \\ &\geq \frac{A}{B}\,\frac{2}{u_{max}\,diam\,K} \end{aligned}$$

The second bound can be shown similarly, but with Corollary 4.2 instead. □

In the grid Example 4.6 the norm is $\|\cdot\|_\infty$, $A = B = u_{max} = 1$ and $diam\,K = k$ so Theorem 4.8 gives a short proof of the edge isoperimetry in this case.

4.3 Rapid Mixing

We are now in a position to apply Average Conductance and Houdré's theorem to obtain bounds on the mixing time and log-Sobolev constants of Markov chains with geometric underlying graphs. The family of Markov chains we will consider is :

Definition 4.2. A *geometric Markov chain* $\mathcal{M}$ is one with a geometric underlying graph $G = (E, V)$, with vertex weights $\pi(v)$ given by the stationary distribution of $\mathcal{M}$ and edge weights $\mathsf{Q}(x, y)$ given by the flow $\pi(x)\,\mathsf{P}(x, y)$ along edges of $\mathcal{M}$.

Example 4.7. Consider the grid $[k]^n$ of Example 4.6. Define a random walk on this graph with equal transition probability $1/(4n)$ to any neighbor. This is a geometric Markov chain with $p = 1/(4n)$, so by (2.1) and (4.17), $\Phi(x) \geq (2 + \log(1/2x))/(4nk)$. Applying the Average Conductance Theorem with $\Phi = \Phi(1/2)$ we get

$$\tau \leq K\left(14\int_{\pi_0}^{1/2} \frac{dx}{x\Phi(x)^2} + \frac{4}{\Phi}\right) \leq 260\,K\,k^2 n^2 = O(k^2 n^2)\ .$$

This is not far from the correct bound of $O(k^2 n \log n)$.

We can extend this technique to general geometric Markov chains in a similar fashion as the extension to edge isoperimetry.

Theorem 4.9. *Let $\mathcal{M}$ be a geometric Markov chain with vertex constant of proportionality A, edge constant B, diameter $diam\, K$ in norm $\|\cdot\|$, and $u(x)$ the Euclidean unit normal to $\partial\phi(v)$ at x satisfying $\|u\|^* \le u_{max}$ for all x along $\partial\phi(v)$. Then*

$$\Phi(x) \ge \frac{2 + \log(1/2x)}{2}\,\Phi_g \qquad \tau \le 38\,K/\Phi_g^2$$

and

$$(i)\ \rho \ge \Phi_g^2/36 \qquad (ii)\ \rho \ge \sqrt{\lambda}\ \Phi_g/24\ .$$

where

$$\Phi_g = \frac{A}{B}\,\frac{2}{u_{max}\, diam\, K}$$

is a lower bound on the conductance (i.e. $\Phi \ge \Phi_g$).

Proof. Let γ be as in Theorem 4.8, then $\Phi_g = \gamma$ and the bound on $\Phi(x)$ follows from Theorem 4.8. The bound on τ follows by substituting this expression and $\Phi \ge \Phi_g$ into the Average Conductance Theorem. The second bound on ρ follows by using the lower bound on $\Phi(x)$ to bound ℓ_1^+ and substituting this into Theorem 2.9. The first bound on ρ comes from substituting $\lambda \ge \Phi^2/2$ into the second bound (or with a weaker constant if Theorem 2.9 (i) is used). □

Observe that Theorem 4.9 can give better mixing time bounds than that given by the spectral gap. One example of this is Example 4.7 where the mixing time was found to be $\tau = O(k^2\, n^2)$, on the other hand the spectral gap is $\lambda = \Omega(1/k^2\, n)$ which would show $\tau = O(k^2\, n^2\, \log k)$, which is a weaker result.

Example 4.8. One Markov chain where geometry has been used to find upper bounds on the mixing time is a random walk on Linear Extensions [DFK91, KK91, Jer98]. Given a partially ordered set $(V, \prec)$, $V = [1..n]$ the set of linear extensions of $\prec$ is defined by

$$\Omega = \{g \in Sym\, V : g(i) \prec g(j) \Rightarrow i \le j, \text{ for all } i, j \in V\}$$

i.e. the set of permutations on V that preserve the partial ordering, or the set of total orderings which are consistent with the partial ordering.

Sample from Ω u.a.r. as follows. If X_t is the current state then choose a transposition $(i, i+1)$ u.a.r., if $X_t \circ (i, i+1) \in \Omega$ then with probability 1/2 set $X_{t+1} = X_t \circ (i, i+1)$, otherwise $X_{t+1} = X_t$. This Markov chain is symmetric so it has the uniform stationary distribution, also all edges (i,j) have identical flow $q_e = 1/(2\,(n-1)|\Omega|)$.

This is a geometric Markov chain as follows. Suppose $g \in \Omega$ is a linear extension. Then let

$$\phi(g) = \{x \in \mathbb{R}^n : 0 \le x_{g(1)} \le x_{g(2)} \le \cdots \le x_{g(n)} \le 1\} \ .$$

To show convexity of $K = \phi(\Omega)$ let $X^0 = \phi(g_0)$ and $X^1 = \phi(g_1)$ be simplexes in K and consider any path $(1-t)a + tb$ from X^0 to X^1. Suppose the path passes between two adjacent simplexes $S^0 = \phi(g_2)$ and $S^1 = \phi(g_3)$, say $x_i \le x_j$ in S^0 but $x_i \ge x_j$ in S^1. Then $x_i - x_j$ can only increase with t (by linearity), in particular X^1 must have $x_i \ge x_j$ and hence the total order defining S^1 was a linear extension, i.e. $S^1 \subseteq K$.

The simplexes all have equal volume and intersections have equal surface area

$$Vol_n(\phi(g)) = \frac{1}{n!} \quad \text{and} \quad Vol_{n-1}(\phi(g_1) \cap \phi(g_2)) = 0 \ or \ \frac{\sqrt{2}}{(n-1)!} \ .$$

The first expression is because each simplex is a $1/n!$ fraction of the cube $[0,1]^n$ of volume 1. The second expression is because the intersection of two simplexes has an $(n-1)$-dimensional projection, e.g. $x_1 \le x_2 \le \cdots \le x_i \le x_{i+2} \le \cdots \le x_n$ of volume $1/(n-1)!$, and the added constraint $x_i = x_{i+1}$ increases this by a factor $\sqrt{2}$.

The stationary and transition probabilities are $\pi(g) = 1/|\Omega|$ and $\mathsf{P}(x,y) = 1/2(n-1)$, so that $A = |\Omega|/n!$ and $B = |\Omega|\,2\sqrt{2}/(n-2)!$. Also, the ℓ_∞ diameter is $diam\,K \le 1$, and the ℓ_1 norms along the boundaries are $\|u\|_1 = \sqrt{2}$. Then Theorem 4.9 gives

$$\begin{aligned}
\Phi_g &= \frac{A}{B}\,\frac{2}{u_{max}\,diam\,K} = \frac{1}{2\sqrt{2}\,n(n-1)}\,\frac{2}{\sqrt{2}} = \frac{1}{2\,n(n-1)} \\
\Phi(x) &\ge \frac{2+\log(1/2x)}{2}\,\Phi_g = \frac{2+\log(1/2x)}{4\,n(n-1)} \\
\tau &\le 152\,K\,n^2(n-1)^2 = O(n^4) \\
\rho &\ge 1/(144n^2(n-1)^2) = \Omega(1/n^4)
\end{aligned}$$

This gives a large improvement over the previous conductance bound of $\tau = O(n^5 \log n)$ and even beats the path-coupling and comparison bound [BD97] of $\tau = O(n^4 \log^2 n)$. It is also quite close to the correct bound of $\Theta(n^3 \log n)$, which Wilson [Wil97] showed by an elegant application of path-coupling.

Chapter 5

Inductive Markov Chains and Balanced Matroids

A second means for bounding the conductance or cutset expansion is by induction. The best example of this in the theory of rapidly mixing Markov chains, is the inductive proof used by Mihail et. al. [MS92, FM92] to bound the cutset expansion of a class of matroids known as balanced matroids. In this section we will extend their results to give an inductive bound on the conductance function $\Phi(x)$ and an improved bound on the mixing time of a Markov chain on balanced matroids. These are results from collaborative work with Jung-Bae Son and previously appeared in [MS01].

5.1 Preliminaries

A *matroid* is an important generalization of objects in many areas of mathematics. There are equivalent definitions of matroids in all these areas, in this section we will follow a form motivated by bases of vector spaces. Much of this description follows an account given in [Jer00].

A *matroid* $\mathcal{M}$ is given by a ground set $E(\mathcal{M})$ and a collection of *bases* $\mathcal{B}(\mathcal{M}) \subseteq 2^{E(\mathcal{M})}$. The bases $\mathcal{B}(\mathcal{M})$ must satisfy two conditions :

- (*Cardinality*) All bases have the same size, namely the *rank* of $\mathcal{M}$.

- (*Edges*) $\forall X, Y \in \mathcal{B}(\mathcal{M}),\ \forall e \in X,\ \exists f \in Y : X \cup \{f\} \setminus \{e\} \in \mathcal{B}(\mathcal{M})$.

The *bases-exchange graph* $G(\mathcal{M})$ of $\mathcal{M}$ has vertex set $\mathcal{B}(\mathcal{M})$, and vertices (bases) are connected by an edge if they differ in exactly one element. The second (edge) condition in the definition of a matroid guarantees that this is a connected graph.

Two operations that reduce the size of matroids are contraction and deletion. Given an element of the ground set $e \in E(\mathcal{M})$, then

- (*Contraction*) $\mathcal{M} \setminus e$ has $E(\mathcal{M} \setminus e) = E(\mathcal{M}) \setminus \{e\}$ and $\mathcal{B}(\mathcal{M} \setminus e) = \{X \subseteq E(\mathcal{M} \setminus e) \mid X \in \mathcal{B}(\mathcal{M})\}$.

- (*Deletion*) $\mathcal{M}/e$ has $E(\mathcal{M}/e) = E(\mathcal{M})\setminus\{e\}$ and $B(\mathcal{M}/e) = \{X \subseteq E(\mathcal{M}/e) \mid X \cup \{e\} \in \mathcal{B}(\mathcal{M})\}$.

Any matroid obtained from a series of contractions and deletions is a *minor* of $\mathcal{M}$.

If X is a basis uniformly chosen at random from $\mathcal{B}(\mathcal{M})$ and e is an element of $E(\mathcal{M})$, let by abuse of notation e denote the event $e \in X$. A matroid $\mathcal{M}$ is *negatively correlated* if for all pairs of distinct elements $e, f \in E(\mathcal{M})$ the inequality $Pr[e|f] \leq Pr[e]$ (eq. $Pr[ef] \leq Pr[e]\, Pr[f]$) holds. A matroid $\mathcal{M}$ is said to be *balanced* if itself and all its minors are negatively correlated.

We define a Markov chain on the bases exchange graph as follows. Suppose the current state is $X \in \mathcal{B}(\mathcal{M})$, then choose a basis element $b \in X$ and an edge $e \in E(\mathcal{M})$ uniformly at random. If $X' = (X \cup e \setminus b) \in \mathcal{B}(\mathcal{M})$ then move to X' with probability $1/2$, otherwise stay at X.

This Markov chain has been shown to mix rapidly by several authors, the strongest bounds on the mixing time were shown in [FM92]. We will apply Average Conductance and log-Sobolev techniques to this problem to obtain improved mixing time bounds.

Remark : A few examples of matroids include :

1. **Graphic matroids :** Ground set is edges of graph, bases are spanning trees, so $m = |E(G)|$ and $n = |V| - 1$.

2. **Vectorial matroids :** Ground set is a vector space, bases are bases of the vector space, so $m = |V|$ and $n = dim\, V$.

5.2 Edge-Isoperimetry

We will first show an edge-isoperimetric inequality for cuts in balanced matroids, just as was done with the geometric problems studied in the previous chapter. The inductive proof is motivated by results in [MS92, FM92].

Theorem 5.1 (Matroid Edge-Isoperimetry). *Let $G(\mathcal{M})$ be the bases-exchange graph of any balanced matroid $\mathcal{M}$ with bases $\mathcal{B}$. For all subsets $\mathcal{S} \subset \mathcal{B}$ with $0 < |\mathcal{S}| \leq |\mathcal{B}|$ we have*

$$\frac{|Cut(\mathcal{S})|}{|\mathcal{S}|} \geq \log_2\left(\frac{|\mathcal{B}|}{|\mathcal{S}|}\right).$$

Proof. The proof will proceed by induction on the size of the ground set of $\mathcal{M}$.

For the base cases, $|E(\mathcal{M})| = 1, 2$, the hypothesis is trivially true.

For the induction step, assume $|E(\mathcal{M})| > 2$. Let $\mathcal{S} \subseteq \mathcal{B}$ define a cut in the bases-exchange graph of $\mathcal{M}$ and let $X = \frac{|\mathcal{S}|}{|\mathcal{B}|}$. Choose any $e \in E(\mathcal{M})$, and let $\mathcal{B}_e = \mathcal{M}/e$ and $\mathcal{B}_{\overline{e}}$ be the deletion and contraction of e respectively. Let $\mathcal{S}_e = \mathcal{S} \cap \mathcal{B}_e$ and $\mathcal{S}_{\overline{e}} = \mathcal{S} \cap \mathcal{B}_{\overline{e}}$, and define x, $y \in [0, 1]$ by $|\mathcal{S}_e| = x\,|\mathcal{B}_e|$ and $|\mathcal{S}_{\overline{e}}| = y|\mathcal{B}_{\overline{e}}|$. Also, define $\alpha \in [0, 1]$ so that $|\mathcal{B}_e| = \alpha|\mathcal{B}|$ and $|\mathcal{B}_{\overline{e}}| = (1-\alpha)|\mathcal{B}|$. Observe that $X = x\,\alpha + y\,(1-\alpha)$.

The edges forming the cut are of three kinds:

1. those whose endpoints are both within $\mathcal{B}_e$
2. those whose endpoints are both within $\mathcal{B}_{\overline{e}}$
3. those which span $\mathcal{B}_e$ and $\mathcal{B}_{\overline{e}}$

Since, as mentioned above, $\mathcal{B}_e$ and $\mathcal{B}_{\overline{e}}$ are isomorphic to $\mathcal{B}(\mathcal{M}/e)$ and $\mathcal{B}(\mathcal{M} \setminus e)$, they give rise to minors of $\mathcal{M}$ and the induction hypothesis is applicable. By the induction hypothesis, the numbers of edges of the first two kinds are at least $x\,\log_2(1/x)\,|\mathcal{B}_e|$ and $y\,\log_2(1/y)\,|\mathcal{B}_{\overline{e}}|$ respectively.

To lower bound the number of edges of type (3), assume first that $x \geq y$. By [FM92](lemma 3.1), there are at least $x|\mathcal{B}_{\overline{e}}|$ bases in $\mathcal{B}_{\overline{e}}$ adjacent to some bases in $\mathcal{S}_e$; of these, at least $(x-y)|\mathcal{B}_{\overline{e}}|$ must lie outside $|\mathcal{S}_{\overline{e}}|$. Thus there are at least $(x-y)|\mathcal{B}_{\overline{e}}|$ edges of type (3).

This argument can equally well be applied in the opposite direction, starting at the set $\mathcal{B}_{\overline{e}} \setminus \mathcal{S}_{\overline{e}}$, yielding a second lower bound of $(x-y)|\mathcal{B}_e|$.

Thus the number of edges of kind (3) is at least $(x - y)\max\{|\mathcal{B}_e|, |\mathcal{B}_{\overline{e}}|\}$. Since the case $x < y$ is entirely symmetric, we obtain, summing the contributions from the three types of edges

$$\begin{aligned} |Cut(\mathcal{S})| &\geq x\log_2\left(\frac{1}{x}\right)|\mathcal{B}_e| + y\log_2\left(\frac{1}{y}\right)|\mathcal{B}_{\overline{e}}| + |x-y|\max\{|\mathcal{B}_e|, |\mathcal{B}_{\overline{e}}|\} \\ &= |\mathcal{B}|\left(\alpha\, x\log_2\left(\frac{1}{x}\right) + (1-\alpha)\, y\log_2\left(\frac{1}{y}\right) + |x-y|\max\{\alpha,\, 1-\alpha\}\right) \end{aligned}$$

To complete the proof, we must show that $|Cut(\mathcal{S})|$ is always at least $X\log_2(1/X)\,|\mathcal{B}|$, where X defined above is $X = \alpha\, x + (1-\alpha)\, y$. It suffices to show

$$\begin{aligned} f(x,y) &= \alpha\, x\log_2\left(\frac{1}{x}\right) + (1-\alpha)\, y\log_2\left(\frac{1}{y}\right) + |x-y|\max\{\alpha,\, 1-\alpha\} - X\log_2(1/X) \\ &\geq 0 \end{aligned}$$

Extend f to the boundary in the obvious way by letting $0\log_2(1/0) = 0$.

Fix α and y, assume $x \neq y$. Then

$$f_x = \frac{\alpha}{\ln 2}\left(\ln\frac{1}{x} - \ln\frac{1}{X}\right) \pm \max\{\alpha, 1-\alpha\}$$

and

$$\begin{aligned} f_{xx} &= \frac{\alpha}{\ln 2}\left(\frac{\alpha}{X} - \frac{1}{x}\right) \\ &= \frac{\alpha}{\ln 2}\frac{-y(1-\alpha)}{x\,X} \\ &\leq 0\,. \end{aligned}$$

Therefore, any extreme points are maxima with respect to x, so a global minimum must occur on the boundary ($x = 0$, $x = 1$, or $x = y$). A similar process shows that $f_{yy} \leq 0$ so $y = 0$, $y = 1$ or $y = x$. This reduces the problem to corners or diagonals. But $f(x,x) = 0$, $f(1,0) = \max\{\alpha, 1-\alpha\} - \alpha\log(1/\alpha) > 0$, and likewise for $f(0,1)$, so the result again holds. □

5.3 Rapid Mixing

We now have all that is needed to show rapid mixing.

Corollary 5.1. *The mixing time of the bases-exchange walk on any balanced matroid of rank n on a ground set of size m is at most* $\tau \leq C\, m^2 n^2$ *for some constant C independent of the matroid.*

Proof. By Theorem 5.1 $\quad |Cut(\mathcal{S})|/|\mathcal{S}| \geq \log_2(1/x) \quad$. The Markov chain has $p = 1/(2mn)$, so by (2.1) in the Preliminaries we have

$$\Phi(x) \geq \frac{1}{2mn} \inf_{\pi_0 \leq \pi(\mathcal{S}) \leq x} \frac{|Cut(\mathcal{S})|}{|\mathcal{S}|} \geq \frac{\log_2(1/x)}{2mn} \tag{5.1}$$

Substituting (5.1) into the Average Conductance theorem gives the result. □

This Theorem is stronger than [FM92] Theorem 5.1 ($\tau = O(n^3 m \log m)$) when $n \log m = \Omega(m)$, eg. when $m = O(n \log n)$. In the case of graphic matroids this would be the case when the average degree of vertices is $O(\log n)$. However, we can get a stronger result with log-Sobolev constants.

Corollary 5.2. *The log-Sobolev constant and mixing time of the bases-exchange walk on any balanced matroid of rank n on a ground set of size m are bounded by*

$$\rho \geq \frac{1}{24\, m^{3/2} n^2} \qquad \tau \leq 24\, m^{3/2} n^2 (\log n + \log\log m)$$

Proof. By (5.1) we see $\ell_1^+ \geq 1/[(2\log 2) m\, n]$. It was shown in [FM92] that $\lambda \geq 1/m\, n^2$. Thus

$$\rho \geq \sqrt{\lambda}\, \ell_1^+/12 \geq 1/(24 m^{3/2}\, n^2)$$

and

$$\tau \leq 12 m^{3/2} n^2 (2 + \log\log(m^n)) \leq 24 m^{3/2} n^2 (\log n + \log\log m)$$

□

This is stronger than [FM92] Theorem 5.1 ($\tau = O(n^3 m \log m)$) when $n \log m = \Omega(\sqrt{m} \log n)$, eg. when $m = O(n^2)$. According to a result by Heller [Hel57] this is true for *simple* regular matroids (i.e. matroids without loops and parallel elements): $m \leq n(n+1)$, which is smaller than $2n^2$ for $n \geq 1$, and implies that $m = O(n^2)$ if the size of all parallel classes is bounded by a constant. In particular, this includes all graphic matroids with few multiple edges.

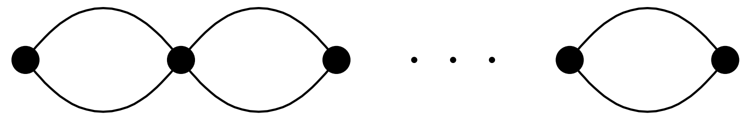

Figure 5.1: A graphic matroid for the binary n-cube

Example 5.1. To see how strong or weak these results are we consider the graphic matroid of Figure 5.1.

This is just the binary n-cube; in the i-th pair of edges the lower edge can be labeled 0 and the upper edge labeled 1 and we let $e_i \in \{0, 1\}$ be the edge that is chosen, then trees are exactly n-tuples $(e_1, e_2, \ldots, e_n)$. The Markov chain on this graphic matroid chooses a coordinate i uniformly and an edge to substitute for it uniformly. With probability $1/n$ the new edge is in the i-th coordinate and we accept the change, otherwise we reject it. Therefore, the spectral gap λ, log-Sobolev constant ρ and mixing time τ are all a factor of n from that of the binary n-cube; this is intuitively clear but can also be proven by using comparison methods of [DSC93].

$$\begin{aligned}
\lambda &= \Theta(1/n^2) &&= \Theta(1/m\,n) \\
\rho &= \Theta(1/n^2) &&= \Theta(1/m\,n) \\
\tau &= \Theta(n^2 \log n) &&= \Theta(m\,n \log n)
\end{aligned}$$

The results in [FM92] showed that $\lambda = \Omega(1/n^3)$ and $\tau = O(n^4 \log n)$, Corollary 5.1 shows $\tau = O(n^4)$, and Corollary 5.2 gives bounds $\rho = \Omega(1/n^{3.5})$ and $\tau = O(n^{3.5} \log n)$.

We note that Jerrum and Son [JS02] have recently shown for balanced matroids that the spectral gap $\lambda \geq 1/(m\,n)$ and log-Sobolev constant $\rho \geq 1/(2\,m\,n)$, so $\tau = O(m\,n\,(\log n + \log\log m))$. On the binary n-cube example shown above, all three bounds are exact up to constant factors. It seems likely that the optimal bound on mixing time are $\tau = O(m\,n \log n)$.

Chapter 6

Canonical Paths and other BCF based Isoperimetric Bounds

This chapter explores applications of the blocking conductance function (BCF) $\phi(x)$, the improvement on Average Conductance which was discussed and proven in Chapter 3.2. In the previous two chapters we considered methods of lower bounding the conductance function, in this chapter we will similarly provide methods for lower bounding the blocking conductance $\phi(x)$.

As a first application we will show how canonical paths can be used to give a blocking conductance function, and thus provide an upper bound on the mixing time. A corollary is closely related to the canonical paths result of Sinclair [Sin92].

Next, for the main result in this chapter we define new conductance-like isoperimetric quantities $h_2^+(x)$ and $h_2^-(x)$. We show that $h_2^+(x)$ can be used to obtain very good upper and lower bounds on the optimal $\phi(x)$, moreover in some cases the quantity $h_2^-(x)$ can be used to further improve the lower bounds.

This method was originally motivated by a theorem of Talagrand [Tal93] where he studied the logarithmic Sobolev constant for the binary n-cube $\{0,1\}^n$, and proved a lower bound for $h_2^+(x)$. Applying Talagrand's bound we obtain a lower bound on $\phi(x)$ and a near optimal mixing time bound for the binary hypercube. As a further application, we show how $h_2^+(x)$ improves on previous conductance based bounds for the mixing time of product Markov chains.

Although the quantity $h_2^+(x)$ suffices to prove near optimal results, the methods given here seem insufficient for obtaining optimal results. We conjecture a stronger mixing time theorem in terms of $h_2^+(x)$. If it holds then it would give optimal mixing time bounds on the binary hypercube, product Markov chains, and most likely also for the geometric and inductive Markov chains considered in the previous two chapters.

It is still possible that the correct lower bound on $\phi(x)$ will give optimal mixing time bounds, however we are unable to show this. The question remains whether the methods used to prove Theorem 3.2 are powerful enough to prove optimal mixing time bounds on problems such as the hypercube. In our second approach to the hypercube we look at the proof of Theorem 3.2 and observe that it only uses subsets of a certain form. We show a blocking conductance function for these subsets, and with Theorem 3.2 this gives the correct mixing time $\tau = O(n \log n)$.

6.1 Canonical Paths

As a first application we show that blocking conductance can be used to give a theorem for canonical paths (see Chapter 2, before Theorem 2.6, for an explanation of canonical paths). We also give a corollary which is closely related to Theorem 2.6.

Theorem 6.1. *Suppose $\mathcal{M}$ is a Markov chain with underlying graph $G = (V, E)$. Define*

$$\rho_e = \max_{e \in E} \frac{1}{\mathsf{Q}(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\, \pi(y) \qquad \textit{and} \qquad \rho_v = \max_{v \in V} \frac{1}{\pi(v)} \sum_{\gamma_{xy} \ni v} \pi(x)\, \pi(y)$$

to be the maximal edge and vertex congestion.

Then

$$\phi(x) = \frac{1}{4\, \rho_v\, \rho_e}$$

is a blocking conductance function, and so

$$\tau = O\left(\rho_v\, \rho_e\, \log \pi_0^{-1}\right) .$$

Proof. In [Sin92] the quantity ρ_e was used to show that $\mathsf{Q}(A, A^c)$ is large. We will use a similar argument but with the quantity ρ_v as well.

Let $A \subset V$ be such that $\pi(A) \leq 1/2$. For each path γ_{xy}, route flow of $\pi(x)\,\pi(y)$ along the path from x to y, for a total of $\pi(A)\,\pi(A^c)$ from A to A^c. Then ρ_v is the highest flow through a vertex as a multiple of its capacity $\pi(v)$.

Let $\partial A = \{x \in A : \mathsf{P}(x, A^c) \neq 0\}$ be the boundary of A with A^c, then

$$\rho_v \geq \frac{\pi(A)\,\pi(A^c)}{\pi(\partial A)}$$

follows because the right side is the average flow per unit of vertex capacity in the boundary, while ρ_v is the worst case.

It follows that a blocking set B of size

$$\lambda = \frac{1}{2}\,\frac{\pi(A)\,\pi(A^c)}{\rho_v} \leq \frac{\pi(A)}{2}$$

will block at most half of the flow $A \to A^c$, so there is at least $\frac{1}{2}\,\pi(A)\,\pi(A^c)$ flow along the edges from $A \setminus B \to A^c$.

Likewise

$$\rho_e \geq \frac{\frac{1}{2}\,\pi(A)\,\pi(A^c)}{\mathsf{Q}(A \setminus B, A^c)}\ ,$$

as the right side is a lower bound on the average flow per unit of edge capacity from $A \setminus B$ to A^c, while ρ_e is the worst case among all vertices.

Then

$$\lambda\,\mathsf{Q}(A \setminus B, A^c) \geq \frac{[\pi(A)\,\pi(A^c)]^2}{4\,\rho_v\,\rho_e}$$

and the result follows from Corollary 3.1. □

The following corollary is similar to the main theorem in [Sin92].

Corollary 6.1. *Let ρ_e and ρ_v be as in Theorem 6.1, and also let*

$$\rho_v^{ave} = \sum_{v \in G} \pi(v) \left[\frac{1}{\pi(v)} \sum_{\gamma_{xy} \ni v} \pi(x)\,\pi(y) \right]$$

be the average vertex congestion over the entire space G. Then

$$\tau = O\left(\frac{\rho_v}{\rho_v^{ave}}\, \rho_e\, \ell_{ave}\, \log \pi_0^{-1} \right)$$

where $\ell_{ave} = \sum_{x,y \in V} \pi(x)\,\pi(y)\,|\gamma_{xy}|$ is the average length of the canonical paths.

Proof. Observe that

$$\begin{aligned} \rho_v^{ave} &= \sum_{\gamma_{xy}} \sum_{v \in \gamma_{xy}} \pi(x)\,\pi(y) \\ &= \sum_{\gamma_{xy}} \pi(x)\,\pi(y)\,(|\gamma_{xy}|+1) \\ &= \ell_{ave}+1 \end{aligned}$$

Multiplying the $\phi(x)$ in Theorem 6.1 by $1 \geq \rho_v^{ave}/2\ell_{ave}$ gives

$$\phi(x) \geq \frac{1}{8}\,\frac{\rho_v^{ave}}{\rho_v}\,\frac{1}{\rho_e\,\ell_{ave}}\ .$$

The result then follows. □

Theorem 6.1 and Corollary 6.1 are essentially equivalent, just written in different forms. Therefore, when comparing the corollary to [Sin92] we will use the simpler Theorem 6.1.

Example 6.1. Consider the balanced matroid problem looked at in Chapter 5 and [FM92]. Results in [FM92] are equivalent to showing $\rho_e \leq \frac{n\,m}{2}$ and $\rho_v \leq 2n$, so by Theorem 6.1 it follows that $\tau = O\left(n^2\,m\,\log\pi_0^{-1}\right)$. This is the same result obtained in [FM92] by using a modified form of [Sin92].

The corollary is not directly comparable to Theorem 2.6. Moreover, it appears that the corollary cannot be weakened further, so that the $\frac{\rho_v}{\rho_v^{ave}}$ term may be necessary. Miclo [Mic99] has shown that for any graph G there is a transition kernel K and distribution μ with $\Phi/\lambda > \ell(G)/2$, where $\ell(G)$ is the length of the longest injective path in G and λ is the spectral gap of the Markov chain given by (G, K). It is easy to construct Markov chains where $\ell_{ave} \ll \ell_{max}$, so this seems to suggest that Sinclair's theorem cannot be weakened to utilize ℓ_{ave} instead of ℓ_{max}, and that Corollary 6.1 must likewise include the $\frac{\rho_v}{\rho_v^{ave}}$ to correct for the use of ℓ_{ave}.

6.2 A new isoperimetric quantity and near optimal bounds on the binary hypercube

The methods of this section are motivated by a desire to apply the following isoperimetric inequality of [Tal93, BG96] to show rapid mixing.

Theorem 6.2 (Talagrand). *For each subset A of $\Omega = \{0,1\}^n$ we have*

$$\frac{\sum_{\alpha\in A}\sqrt{P(\alpha,\, A^c)}\,\pi(\alpha)}{\pi(A)\,\pi(A^c)} \geq \frac{1}{4}\sqrt{\frac{\log\frac{1}{\pi(A)\,\pi(A^c)}}{n}}\ .$$

This suggests a conductance type function :

Conductance	:	$\tilde{\Phi}$	$=$	$\min_{0<\pi(A)\leq 1/2}$	$\frac{\sum_{\alpha\in A}\pi(\alpha)\,\mathsf{P}(\alpha,A^c)}{\pi(A)\,\pi(A^c)}$
Conductance function	:	$\tilde{\Phi}(x)$	$=$	$\min_{0<\pi(A)\leq x}$	$\frac{\sum_{\alpha\in A}\pi(\alpha)\,\mathsf{P}(\alpha,A^c)}{\pi(A)\,\pi(A^c)}$
h_2	:	$h_2^+(x)$	$=$	$\min_{0<\pi(A)\leq x}$	$\frac{\sum_{\alpha\in A}\pi(\alpha)\,\sqrt{\mathsf{P}(\alpha,A^c)}}{\pi(A)\,\pi(A^c)}$

Some history of these quantities serves to further motivate $h_2^+(x)$. Houdré and Tetali [HT96] defined a quantity h_2^+, which is just $h_2^+(1/2)$, as part of a larger family

$$h_p^+ = \min_{0<\pi(A)\leq 1/2}\frac{\sum_{\alpha\in A}\pi(\alpha)\,\sqrt[p]{\mathsf{P}(\alpha,A^c)}}{\pi(A)\,\pi(A^c)}\ .$$

(their quantity was slightly different, actually a factor of 1 to 2 smaller than our h_p^+). The quantity h_1^+ is the Cheeger constant h and roughly the same as Φ, and relates to edge isoperimetry, while h_∞^+ is related to vertex isoperimetry [BHT00]. Both h_1^+ and h_∞^+ can be used to show rapid mixing by bounding the spectral gap λ [BHT00]. Talagrand [Tal93] observed that a simple application of Cauchy-Shwartz gives the upper bound

$$h_2^+(x) \leq \frac{\sqrt{\pi(\partial_{int}A)\,\mathsf{Q}(A,A^c)}}{x\,(1-x)}\ ,$$

which is easily seen to also be an upper bound on $\sqrt{\phi(x)}$.

A similar family h_p^- and $h_p^-(x)$ can also be defined by replacing $\mathsf{P}(\alpha,\,A^c)$ for $\alpha\in A$ with $\mathsf{P}(\alpha,\,A)$ for $\alpha\in A^c$; these relate to exterior vertices and all the above comments hold with h_p^- in place of h_p^+.

We could also define a $\Phi^+(x)$ and $\Phi^-(x)$, just as was done in the definitions of $h_2^\pm(x)$, but by time reversibility we would have $\Phi^+(x) = \Phi^-(x)$. On the other hand, $h_2^+(x)$ and $h_2^-(x)$ can be quite different [Tal93]. Therefore, we may also consider a symmetric version

$$h_2(x) = \min_{0<\pi(A)\leq x}\frac{\sum_{\alpha\in A}\pi(\alpha)\,\sqrt{\mathsf{P}(\alpha,A^c)}+\sum_{\alpha\in A^c}\pi(\alpha)\,\sqrt{\mathsf{P}(\alpha,A)}}{\pi(A)\,\pi(A^c)}\ .$$

This is similar to a quantity h_2 in [HT96].

Also, let $h_2^+(A)$ be defined as would be expected, by

$$h_p^+(A) = \frac{\sum_{\alpha\in A}\pi(\alpha)\,\sqrt[p]{\mathsf{P}(\alpha,A^c)}}{\pi(A)\,\pi(A^c)}\ ,$$

and likewise for $h_p^-(A)$ and $h_p(A)$.

Our main result in this section will be the following theorem.

Theorem 6.3. *Given a Markov chain $\mathcal{M}$ with state space G, let $\mathsf{P}_{min} = \min_{x,y\in G}\{\mathsf{P}(x,y) : \mathsf{P}(x,y) > 0\}$. Then for any set $A \subset G$ with $\pi(A) \le \frac{1}{2}$ it follows that*

$$\begin{aligned} 2\,h_2^+(A)^2 &\ge \Psi_{int}(A) \ge \left(\frac{h_2^+(A)}{2+\log(1/\sqrt{\mathsf{P}_{min}})}\right)^2 \\ \min\{2\,h_2^-(A)^2,\, 4\Phi(A)\} &\ge \Psi_{ext}(A) \ge \frac{1}{4}\sqrt{\mathsf{P}_{min}}\,h_2^-(A)\,\min\{2,\, h_2^-(A)\}\ . \end{aligned}$$

Corollary 6.2. *The mixing time of the lazy random walk on the cube $\{0,1\}^n$ is*

$$\tau = O\left(n\,\log^3 n\right)\ .$$

Proof of Corollary 6.2. Use the previous theorem with $h_2^+(x)$ and $\hat{\phi}(x) = \min_{\pi(A)=x}\Psi_{int}(A)$. Observe that Talagrand's bound on $h_2^+(x)^2$ varies by at most a factor of 2 when $y \in [x/2, x]$, so it suffices to let $\phi(x) = \frac{1}{4}\,\hat{\phi}(x)$. □

In the proofs of all the bounds we let $v_1,\, v_2,\, \ldots,\, v_k$ be the elements of A, assume that they are ordered in decreasing order of $\mathsf{P}(v_i,\, A^c) = \mathsf{Q}(v_i,\, A^c)/\pi(v_i)$, and let $B_i = \{v_1,\, v_2,\, \ldots,\, v_i\}$ ($B_0 = \emptyset$ is the empty-set).

The upper bounds in the theorem are the easier to prove of the two directions.

Proof of upper bounds. Looking back at the definition of $\Psi_{int}(A)$, observe that upper bounding $\Psi_{int}(A)$ requires only an upper bound on

$$[\pi(B_i) + \pi(v_{i+1})]\,\mathsf{Q}(A\setminus B_i,\, A^c) = \pi(B_{i+1})\,\mathsf{Q}(A\setminus B_{i+1},\, A^c) + \pi(B_{i+1})\,\mathsf{Q}(v_{i+1},\, A^c) \quad (6.1)$$

for all i. We will bound the two terms on the right separately.

Let $B = B_\ell$ be the "maximal blocking set" in the sense that

$$\forall i,\, \pi(B_i) \le \frac{1}{2} : \pi(B_i)\,\mathsf{Q}(A\setminus B_i,\, A^c) \le \pi(B)\,\mathsf{Q}(A\setminus B,\, A^c)\ .$$

In particular, if $y = v_{\ell+1}$ then

$$[\pi(B) + \pi(v_{\ell+1})]\,[\mathsf{Q}(A\setminus B,\, A^c) - \mathsf{Q}(v_{\ell+1},\, A^c)] \le \pi(B)\,\mathsf{Q}(A\setminus B,\, A^c)\ .$$

Simplifying the inequality gives

$$\mathsf{P}(v_{\ell+1},\, A^c) = \frac{\mathsf{Q}(v_{\ell+1}, A^c)}{\pi(v_{\ell+1})} \geq \frac{\mathsf{Q}(A \setminus B,\, A^c)}{\pi(B) + \pi(v_{\ell+1})} \ .$$

Because the $v_i \in B$ are in decreasing order then $\mathsf{P}(v_i,\, A^c) \geq \mathsf{P}(v_{\ell+1},\, A^c)$, and therefore

$$\begin{aligned} h_2^+(A)\,\pi(A)\pi(A^c) &= \sum_{\alpha \in A} \pi(\alpha)\sqrt{\mathsf{P}(\alpha,\, A^c)} \\ &\geq \sum_{\alpha \in B \cup y} \pi(\alpha)\sqrt{\mathsf{P}(v_{\ell+1},\, A^c)} \\ &\geq \sum_{\alpha \in B \cup y} \pi(\alpha)\sqrt{\frac{\mathsf{Q}(A \setminus B,\, A^c)}{\pi(B) + \pi(v_{\ell+1})}} \\ &= \sqrt{[\pi(B) + \pi(v_{\ell+1})]\,\mathsf{Q}(A \setminus B,\, A^c)} \ . \end{aligned}$$

This gives a bound on the first term on the right of (6.1).

For the second term, observe that

$$\begin{aligned} h_2^+(A)\,\pi(A)\pi(A^c) &\geq \sum_{j \leq i+1} \pi(v_j)\,\sqrt{\mathsf{P}(v_{i+1},\, A^c)} \\ &= \pi(B_{i+1})\,\sqrt{\mathsf{P}(v_{i+1},\, A^c)} \\ &\geq \sqrt{\pi(B_{i+1})\,\mathsf{Q}(v_{i+1},\, A^c)} \end{aligned}$$

It follows that

$$\Psi_{int}(A) \leq 2\, h_2^+(A)^2 \ .$$

The same proof will hold with $\Psi_{ext}(A)$ and $h_2^-(A)$. The $4\Phi(A)$ bound comes from $\pi(A) \geq \pi(B)$ and $\mathsf{Q}(A, A^c) \geq \mathsf{Q}(A, A^c \setminus B)$, so that $\Psi_{ext}(A) \leq \pi(A)\,\mathsf{Q}(A, A^c)/(\pi(A)\pi(A^c))^2 \leq 4\,\Phi(A)$. □

The proof of the upper bound treated $\Psi_{int}(A)$ as a bound on $\pi(B)\,\mathsf{Q}(A \setminus B,\, A^c)$, with the $v_{\ell+1}$ term merely complicating the proof. In contrast to this, for both lower bounds the $v_{\ell+1}$ term will prove essential.

A first approach to a lower bound is to bound the external (or internal) vertex boundary, and then bound edge isoperimetry when a constant fraction of these vertices are blocked. This was the approach used with canonical paths in the previous section, and in proving the second lower bound.

Proof of lower bound on $\Psi_{ext}(A)$. Define

$$\mathsf{Q}_2(C,D) = \sum_{\alpha\in C} \pi(\alpha)\,\sqrt{\mathsf{P}(\alpha,D)}\ .$$

Suppose that $B' \subset A^c$ is such that $\mathsf{Q}_2(B',A) > \frac{1}{2}\,\mathsf{Q}_2(A^c,A)$. Because $\mathsf{P}(A^c,\cdot) \le 1$ then

$$\frac{1}{2}\,\mathsf{Q}_2(A^c,A) < \mathsf{Q}_2(B',A) \le \pi(B')\ .$$

It follows that if $\lambda = \min\{\frac{1}{2}\,\mathsf{Q}_2(A,A^c),\,\pi(A)\}$ then $\mathsf{Q}_2(A^c\setminus B,A) \ge \frac{1}{2}\,\mathsf{Q}_2(A^c,A)$. Therefore,

$$\begin{aligned}
\lambda\,\mathsf{Q}(A^c\setminus B,A) &\ge \min\left\{\frac{1}{2}\,\mathsf{Q}_2(A^c,\,A),\,\pi(A)\right\}\,\mathsf{Q}_2(A^c\setminus B,A)\,\sqrt{\mathsf{P}_{min}} \\
&\ge \frac{1}{2}\,\mathsf{Q}_2(A^c,A)\,\min\left\{\frac{1}{2}\,\mathsf{Q}_2(A^c,\,A),\,\pi(A)\right\}\,\sqrt{\mathsf{P}_{min}}\ .
\end{aligned}$$

Essentially the same proof applies to the $\Psi'_{ext}(A)$ discussed in the remarks following Theorem 3.2. With $\Psi'_{ext}(A)$ we can allow $\lambda = \frac{1}{2}\,\mathsf{Q}_2(A^c,A)$. □

This is already a major improvement over previous isoperimetric methods, for example for the lazy Markov chain on the binary n-cube this gives mixing in time $\tau = O(n^{3/2}\,\log n)$, while $\tau = O(n^3)$ by conductance or $\tau = O(n^2)$ by average conductance. However, by more carefully bounding the blocking conductance we are able to improve on this.

In order to prove the lower bound on $\Psi_{int}(A)$ we will treat the set of points $A = \{v_1,\,v_2,\,\ldots,\,v_k\}$ as a continuous set $[0,x]$ where $x = \pi(A)$, and treat $\mathsf{P}(v_i,\,A^c)$ as a continuous non-increasing function on $A = [0,x]$ (recall that the v_i are ordered by decreasing $\mathsf{P}(\cdot,\,A^c)$). The probability measure π and the ergodic flow $\mathsf{Q}(\cdot,\,A^c)$ can then be naturally extended to $\pi([a,b]) = \int_a^b 1\,dt = b-a$ and $\mathsf{Q}([a,b],\,A^c) = \int_a^b \mathsf{P}(t,A^c)\,dt$. The "blocking sets" will be $B = [0,b] \subset A$, and we also define

$$\Psi_0(A) = \sup_{\lambda\le x}\ \inf_{B=[0,b]\subseteq[0,\lambda]} \lambda\,\mathsf{Q}(A\setminus B,\,A^c) = \sup_{\lambda\le x} \lambda\,\mathsf{Q}([\lambda,x],\,A^c)\ .$$

With a little thought it should be clear that $\Psi_{int}(A) \ge \Psi_0(A) \ge \frac{1}{4}\,\Psi_{int}(A)$, so in order to lower bound $\Psi_{int}(A)$ it will suffice to lower bound $\Psi_0(A)$. In the following proof we will always work with intervals $B = [0,b]$ and $A\setminus B = [b,x]$.

Proof of lower bound on $\Psi_{int}(A)$. Define $B_0 = [0,\lambda]$ to be the B where the optimal value $\Psi_0(A)$ occurs. Then for any other $B = [0,b]$ it follows that

$$\pi([0,b])\,\mathsf{Q}([b,x],\,A^c) \le \pi(B_0)\,\mathsf{Q}(A\setminus B_0,\,A^c) = \Psi_0(A)\ ,$$

and therefore,

$$\mathsf{Q}([b,x],\, A^c) \leq \frac{\Psi_0(A)}{\pi(B)} \ . \tag{6.2}$$

If $B = [0, \sqrt{\Psi_0(A)/\mathsf{P}_{min}}]$ then this shows that $\mathsf{Q}([\sqrt{\Psi_0(A)/\mathsf{P}_{min}},\, x],\, A^c) \leq \sqrt{\Psi_0(A)\,\mathsf{P}_{min}}$. When $\mathsf{P}(t,\, A^c) \neq 0$ then $\mathsf{P}(t,\, A^c) \geq \mathsf{P}_{min}$ (by definition of P_{min}), so this flow can be distributed over a set of size at most $\sqrt{\Psi_0(A)\,\mathsf{P}_{min}}/\mathsf{P}_{min}$. It follows that $\mathsf{P}(t, A^c) = 0$ if $t > 2\sqrt{\Psi_0(A)/\mathsf{P}_{min}}$. With this in mind, then (6.2) can be rewritten as an integral

$$\int_b^x \mathsf{P}(t,\, A^c)\, dt \leq \int_b^x \mathsf{P}_c(t)\, dt$$

where

$$\mathsf{P}_c(t) = \begin{cases} 1 & if\ t \leq \sqrt{\Psi_0(A)} \\ \Psi_0(A)/t^2 & if\ \sqrt{\Psi_0(A)} < t < \sqrt{\Psi_0(A)/\mathsf{P}_{min}} \\ \mathsf{P}_{min} & if\ \sqrt{\Psi_0(A)/\mathsf{P}_{min}} \leq t \leq 2\sqrt{\Psi_0(A)/\mathsf{P}_{min}} \\ 0 & if\ t > 2\sqrt{\Psi_0(A)/\mathsf{P}_{min}} \end{cases}$$

By Lemma 6.1 (see below) it follows that

$$\int_0^x \sqrt{\mathsf{P}(t,\, A^c)}\, dt \leq \int_0^x \sqrt{\mathsf{P}_c(t)}\, dt \ .$$

Rewriting the left side and integrating the right side shows that

$$h_2^+(A)\, \pi(A)\pi(A^c) = \int_0^x \sqrt{\mathsf{P}(t,\, A^c)}\, dt \leq \sqrt{\Psi_0(A)}(2 + \log(1/\sqrt{\mathsf{P}_{min}})) \ .$$

□

Lemma 6.1. *Suppose that $f,\, g : \mathbb{R}^+ \to \mathbb{R}^+$, and that g is monotonically increasing. If*

$$\forall t \geq 0 : \int_0^t f(y)\, dy \leq \int_0^t g(y)\, dy$$

then it follows that

$$\forall t \geq 0 : \int_0^t \sqrt{f(y)}\, dy \leq \int_0^t \sqrt{g(y)}\, dy \ .$$

Proof of Lemma. We prove the lemma when the set $S = \{y : f(y) > g(y)\}$ can be decomposed into a finite number of disjoint intervals S_i, as this is sufficient for our proof of Theorem 6.3.

Assume that the S_i are ordered so that if $i < j$ then $S_i < S_j$ (i.e. if $x \in S_i$ and $y \in S_j$ then $x < y$). By the assumption in the theorem, if $\int_{S_1} (f(y) - g(y))\, dy = M_1$ then there exists a set $T_1 < S_1$ such that $\int_{T_1} (g(y) - f(y))\, dy = M_1$. Likewise, if $\int_{T_2} (f(y) - g(y))\, dy = M_2$ then there exists a set $T_2 < S_2$ disjoint from $T_1 \cup S_1$ and such that $\int_{T_2} (g(y) - f(y))\, dy = M_2$. This can be repeated for all S_i, generating T_i such that $T_i < S_i$ and all S_i and T_i are disjoint. Also, let x_i be the left endpoints of the S_i.

Then

$$\begin{aligned}
\int_0^t \left(\sqrt{g(y)} - \sqrt{f(y)}\right) dy &= \int_0^t \frac{g(y) - f(y)}{\sqrt{g(y)} + \sqrt{f(y)}}\, dy \\
&\geq \int_0^t \frac{g(y) - f(y)}{2\sqrt{g(y)}}\, dy \\
&\geq \sum_i \int_{S_i} \frac{g(y) - f(y)}{2\sqrt{g(y)}}\, dy + \int_{T_i} \frac{g(y) - f(y)}{2\sqrt{g(y)}}\, dy \\
&\geq \sum_i \int_{S_i} \frac{g(y) - f(y)}{2\sqrt{g(x_i)}}\, dy + \int_{T_i} \frac{g(y) - f(y)}{2\sqrt{g(x_i)}}\, dy \\
&= 0\,.
\end{aligned}$$

□

This completes the proof of Theorem 6.3.

6.3 Examples

Perhaps the best example of how Theorem 6.3 improves on previous geometric bounds on mixing time is the lazy random walk on the binary hypercube $\{0,1\}^n$, where it was shown in Corollary 6.2 that $\tau = O(n \log^3 n)$. In the following we give other applications.

Example 6.2. Consider the natural Markov chain on the complete graph K_n given by choosing among neighboring vertices with probability $\frac{1}{n-1}$ and holding with probability 1/2. The following quantities are not difficult to compute :

- $\tau = O(1)$,
- $\Psi_{int}(x) = \frac{1}{8(1-x)}$, $\Psi_{ext}(x) = \frac{1}{2(1-x)}$,
- $h_2^+(x) = \frac{1}{\sqrt{2(1-x)}}$, $h_2^-(x) = \frac{1}{\sqrt{2x}}$, $\Phi(x) = \frac{1-x}{2}$, $\mathsf{P}_{min} = \frac{1}{2n}$,

where we let $\Psi_{int}(x) = \inf_{\pi(A)=x} \Psi_{int}(A)$, and likewise for $\Psi_{ext}(x)$

Then Theorem 6.3 for $h_2^+(x)$ gives the bound $1 \geq \Psi_{int}(x) \geq \frac{1}{2(1-x)(2+\log\sqrt{2n-2})^2}$, and mixing time $\tau = O(\log^3 n)$. However, the bound involving $h_2^-(x)$ gives $2(1-x) \geq \Psi_{ext}(x) \geq 1/4\sqrt{x(n-1)}$ and mixing in time $\tau = O(\sqrt{n})$.

Another natural Markov chain to consider is the weighted two point space, with weights p and $q = 1-p$, and the natural lazy random walk on this graph with transition probabilities from p to q of $1/2q$ and from q to p of $1/2p$. Then, for $p \ll q$ we have

- $\tau = O(1)$,
- $\Psi_{int}(x) = \frac{1}{2q}$, $\Psi_{ext}(x) = \frac{1}{2q}$,
- $h_2^+(x) = \frac{1}{\sqrt{2q}}$, $h_2^-(x) = \frac{1}{\sqrt{2p}}$, $\Phi(x) = \frac{q}{2}$, $\mathsf{P}_{min} = \frac{p}{2}$.

Then Theorem 6.3 for $h_2^+(x)$ gives the bound $1/q \geq \Psi_{int}(x) \geq 1/2 \log^2(1/p)$ and mixing in time $\tau = O(\log^3(1/p))$. However, the bound in terms of $h_2^-(x)$ gives $2q \geq \Psi_{ext}(x) \geq 1/4$, and mixing in time $\tau = O(\log(1/p))$. The bound on τ in terms of $h_2^-(x)$ can be made optimal ($\tau = O(1)$) by observing that we can let $\phi(x) = \infty$ when $x > 2\,p$.

These examples show that either of the lower bound on $\Psi_{ext}(A)$ or $\Psi_{int}(A)$ may be better for bounding mixing time, depending on the problem : the $h_2^+(x)$ lower bound for K_n and the $h_2^-(x)$ bound for the two point space. In contrast, with the binary hypercube considered in the previous section, as well as the complete graph and two point space just considered, the upper bounds for $\Psi_{int}(A)$ and $\Psi_{ext}(A)$ are all within constant factors of the correct values.

Unfortunately, the constructions in the proof of the lower bounds show that the log terms cannot be dropped in general.

Example 6.3. Consider a Markov chain $\mathcal{M}$ as before.

Houdré and Tetali [HT96] showed that for a product Markov chain $K^n = K_1 \times K_2 \times \cdots \times K_n$ that

$$h_p^+(K^n) \geq \frac{1}{4\sqrt{6}\,n^{1/p}} \min_{1\leq i\leq n} h_1^+(K_i) \qquad \mathit{for} \quad 1 \leq p \leq 2 \; . \tag{6.3}$$

Setting $p = 2$ this gives

$$h_2^+(K^n) \geq \frac{1}{4\sqrt{6}\,\sqrt{n}} \min_{1\leq i\leq n} \Phi(K_i) \; .$$

Observe that $h_2^+(x) \geq h_2^+(K^n)$.

Then applying Theorem 6.3 shows

$$\tau(K^n) = O\left(n\left(\log n + \log(1/\min_{1\leq i\leq n} \mathsf{P}_{i,\,min})\right)^2 \frac{\log(1/\pi_0)}{\min_{1\leq i\leq n} \Phi(K_i)^2}\right) .$$

Previous geometric proofs for mixing time were only able to show

$$\tau(K^n) = O\left(n^2 \frac{\log(1/\pi_0)}{\min_{1\leq i\leq n} \Phi(K_i)^2}\right) ,$$

and so working with h_2^+ will typically improve bounds by a factor of $n/\log^2 n$.

This is also nearly as strong as the bound using the spectral gap. Recall from Theorem 2.5 that $\lambda(K_i) \geq 1/\Phi(K_i)^2$. Using the well known fact that $\lambda(K^n) = \min\{\lambda(K_i)\}/n$ gives $\lambda(K^n) \geq 1/n\ \min_{1\leq i\leq n} \Phi(K_i)^2$, and therefore a slightly better mixing time of

$$\tau(K^n) = O\left(n \frac{\log(1/\pi_0)}{\min_{1\leq i\leq n} \Phi(K_i)^2}\right) .$$

Example 6.4. Tensorizing h_2^+ does not exploit the full power of blocking conductance, because it does not consider the sizes of sets, as does $h_2^+(x)$. Houdré [Hou01] introduced the quantity g_p^+ which also considers the sizes of sets via a log factor,

$$g_p^+(K)^2 = \inf_{\substack{A\subset K \\ \pi(A)\leq 1/2}} \frac{\left(E_\pi D_p^+ 1_A\right)^2}{\pi(A)\pi(A^c)\sqrt{\log(1/\pi(A)\pi(A^c))}}$$

where

$$D_p^+ f(x) = \left(\sum_{y\in K} \mathsf{P}(x,y)[f(x)-f(y)]^{+p}\right)^{1/p}$$

is the discrete p-gradient at x, 1_A is the indicator function of A and $C^+ = \max\{C, 0\}$ is the positive component of C.

Applying Theorem 6.3 to this shows

$$\tau(K) = O\left(\frac{1}{g_2^{+2}} \log^2(1/\mathsf{P}_{min}) \log\log(1/\pi_0)\right) , \tag{6.4}$$

so that g_2^{+2} is nearly as strong as the log-Sobolev constant. This is a new result, although Houdré [Hou01] has shown a relation between g_1^+ and the log-Sobolev constant. It is not yet known how to bound g_2^+ for more than a few spaces, but it may turn out to be easier than bounding the logarithmic Sobolev constant.

We give here a lower bound on g_2^+ for the random walk on the grid $[k]^n$ considered in Chapter 4. This will lead to a mixing time bound close to the correct $\tau = \Theta(k^2\, n\, \log n)$.

Bobkov's constant b_p^+ is defined to be the largest constant such that for all $f : X \to [0,1]$,

$$I_{gauss}(Ef) \leq E\sqrt{I_{gauss}^2(f) + (D_p^+ f)^2 / b_p^{+2}} \ ,$$

where I_{gauss} is the Gaussian Isoperimetric constant considered in Example 4.3. Recall that $I_{gauss}(0) = I_{gauss}(1) = 0$. Moreover, since $I_{gauss}(x) \geq x(1-x)\sqrt{\log(1/x(1-x))}$ ([BG96]) this implies that when $f = 1_A$ is an indicator function then $g_p^+ \geq b_p^+$.

It is well known (see for example [Mur01]) that Bobkov's constant b_2^+ tensorizes as

$$b_2^+(K^n) = \frac{1}{\sqrt{n}}\, b_2^+(K) \ .$$

Lower bounding $b_2^+(K)$ is difficult. A weaker result can be found by using $b_2^+(K) \geq b_1^+(K)$, which follows trivially from the definitions. In her Ph.D. Dissertation Murali [Mur01] has shown that the quantity

$$\beta_1^{+2}(K) = \min_{\substack{A \subset K \\ \pi(A) \leq 1/2}} \frac{\Phi^2(K)}{\pi(A)\, \pi(A^c)}$$

satisfies $b_1^+(K) \geq \beta_1^+(K)$. We then have the following chain of inequalities

$$g_2^+(K^n) \geq b_2^+(K^n) = \frac{1}{\sqrt{n}}\, b_2^+(K) \geq \frac{1}{\sqrt{n}}\, b_1^+(K) \geq \frac{1}{\sqrt{n}}\, \beta_1^+(K) \ ,$$

or in particular

$$h_2^+(x) \geq \frac{1}{\sqrt{n}}\, \beta_1^+(K)\, \log(1/x) \ .$$

It is easy to bound $\beta_1^+(K)$. In particular, for the natural lazy random walk on the line $[k]$, it is clear that $\Phi(x) = \Theta(1/kx)$, and so $\beta_1^+(K) = \Theta(1/k^2)$. Applying (6.4) to this shows

$$\tau([k]^n) = O\left(k^2\, n\, \log^3 n\right) \ . \tag{6.5}$$

Remark : In both the hypercube example and the direct product example, only the $\log^2(1/\mathsf{P}_{min})$ term kept the bounds from being correct. A method to bound the mixing time in terms of $h_2^+(x)$ might give mixing in

$$\tau = O\left(\int_{\pi_0}^{\frac{1}{2}+\pi_{max}} \frac{dx}{x\, h_2^+(x)^2}\right) \ ,$$

where $\pi_0 = \min \pi(v)$ and $\pi_{max} = \max \pi(v)$. This would give correct mixing time results on a variety of Markov chains such as the line, binary hypercube, dumbbell, and products of Markov chains. Also, for the geometric and inductive Markov chains considered in the previous chapters this would likely give correct mixing time bounds (see Chapter 7). Moreover, Theorems 2.1 and 2.2 – the mixing time bounds for conductance and average conductance – would follow immediately because $h_2^+(x) \geq \Phi(x)$.

6.4 Optimal bounds with a modified blocking conductance

In Section 6.2 blocking conductance was used to prove that the lazy random walk on the binary hypercube $\{0,1\}^n$ is mixing in time $\tau = O(n \log^3 n)$. Although this is near optimal, there is still a gap of size $\log^2 n$ between this bound and the correct $\tau = \Theta(n \log n)$. This gap is the same size as the gap between our upper and lower bounds on Ψ_{int}, so it is quite possible that blocking conductance can be used to prove the correct mixing time for this Markov chain. At the moment we have been unable to show this either in the affirmative or in the negative, but in this section we are able to show that a slightly weaker version of blocking conductance can be used to prove the correct bound for this Markov chain. Our main tool will be a compression method that builds on ideas of Bollobas and Leader.

Bollobas and Leader considered the binary hypercube (actually, the grid $[k]^n$) and showed that compressing sets A along the axes e_i reduces the numbers of cut edges [BL91b] and neighboring vertices [BL91a]. A similar argument, by compressing A and B simultaneously, can be used to show that compression of a set A leads to a set A' with smaller blocking conductance. Repeated compression along the various axes will lead to a down-set $\mathcal{A}$ with small blocking conductance. However, it is unclear how to reduce $\mathcal{A}$ to a canonical form, as was done in the Bollobas and Leader results (lexicographic order [BL91b] and simplicial order [BL91a]).

To see where the difficulty arises, consider the case when $x = 1/2$ and n is odd and use exterior blocking conductance ($\Psi_{ext}(A)$). The set A with the fewest cut edges is $\mathcal{A}_e = \{x : x_n = 0\}$, and this set will have $\lambda = \frac{1}{4}$ or equivalently there will be 2^{n-2} vertices in the blocking set B. The set with the least neighboring vertices is $\mathcal{A}_v = \{x : \sum_{i=1}^n x_i \leq \frac{n-1}{2}\}$, and it's blocking set B will have $\frac{1}{2}\binom{n}{(n+1)/2} \ll 2^{n-2}$ vertices in the blocking set. Therefore, any compression method will need to not only compress from $\mathcal{A}_v$ to $\mathcal{A}_e$ (or vice-versa), but it will also need to adjust the sizes of the blocking sets B. It seems difficult to construct such a compression.

This motivates a different approach to Blocking Conductance. In this section we look at the proof of Theorem 3.2 and observe that the blocking conductance only needs to be bounded for sets $\mathcal{A}$ of a certain form. For these sets we are able to compress "diagonally," leading to a fractional hamming ball as the worst case BCF.

In the search for an optimal BCF we delve into the proof of Theorem 3.2.

Lemma 6.2. *Suppose the space G can be decomposed into a disjoint sum $G = \coprod_{i\in\mathcal{I}} S_i$ where $\forall x, y \in S_i \ : \ g(x) = g(y)$. Then it suffices to show a BCF for $\mathcal{I}', \mathcal{I}'' \subset \mathcal{I}$ disjoint ($\mathcal{I}' \cap \mathcal{I}'' = \emptyset$), $A = \coprod_{i\in\mathcal{I}'} S_i$ and $B = \coprod_{i\in\mathcal{I}''} S_i$.*

More simply, it suffices to bound $\phi(x)$ only for A and B disjoint unions of S_i.

Proof. The proof of Theorem 3.2 uses $\phi(\cdot)$ to bound $Q(A, C)$ for some disjoint sets A and C. The goal was to bound $g(v_i) - g(v_j)$ where $v_i \in A$ has minimal $g(v_i)$ in A and $v_j \in C$ has maximal $g(v_j)$ in C. But then, if $v_i \in S_k$ then $\forall v' \in S_k : g(v') = g(v_i)$, so in particular we can assume that $S_k \subseteq A$ and this does not effect the value of the smallest $g(v_i)$ in A. Likewise, if $v_j \in S'_k$ then we can assume $S'_k \subseteq C$. □

Corollary 6.3. *Bounding blocking conductance for the binary hypercube reduces to considering the case when A and B are disjoint unions of $S_k = \{x : \sum_{i=1}^n x_i = k\}$.*

Proof. It is easy to see that the mixing time is bounded by the time from the worst starting point.

The hypercube 2^n is vertex transitive, so up to automorphism the distribution does not depend on the starting point. Without loss assume the starting point is the origin 0.

We can define level sets $S_k = \{x : d(0, x) = k\}$. This follows from distance transitivity; if $d(0, x) = d(0, y) = k$ and g is an automorphism such that $g(0) = 0$, $g(x) = y$, then unique arcs of length t from 0 to x get mapped to unique arcs of length t from 0 to y, and conversely for g^{-1}, so $\mathsf{p}^{(t)}(x) = \mathsf{p}^{(t)}(y)$ for all $t \geq 0$. □

To define our compression operators on these S_k we will need some additional notation. Observe that a set $A \subset G$ can be represented by an indicator function $\mathsf{1}_A$, and that we can define $\pi(A)$, $\mathsf{Q}(A, A^c)$, etc. in terms of $\mathsf{1}_A$, for example $\mathsf{Q}(A, A^c) = \sum_{x \in A,\, y \in A^c} (\mathsf{1}_A(x) - \mathsf{1}_A(y))^+ \, \pi(x) \mathsf{P}(x, y)$. The definitions below will build on this by generalizing the indicator $\mathsf{1}_A(\cdot) \in \{0, 1\}$ to a probability $P_A(\cdot) \in [0, 1]$.

A *fractional set* A is given by $P_A(x) \in [0, 1]$, the probability that $x \in A$. If π is a distribution on G then we can define $\pi(A)$ in terms of $P_A(\cdot)$, as well as cut edges and boundary vertices, as given below. Other definitions, such as those in [BL91c] may be suitable for other situations.

$$\begin{aligned}
\pi(A) &= \sum_{x \in G} P_A(x)\, \pi(x) \\
P_{\partial A}(x) &= max_{y \in \Gamma(x)} \left(P_A(y) - P_A(x)\right)^+ \\
\mathsf{Q}(\vec{e} = \{y, x\}) &= \left(P_A(y) - P_A(x)\right)^+ \\
\pi(\partial A) &= \sum_{x \in G} P_{\partial A}(x)\, \pi(x) \\
\mathsf{Q}(A, A^c) &= \sum_{\vec{e} = \{y,x\} \in G \times G} \mathsf{Q}(\vec{e})\, \pi(y)\, \mathsf{P}(y, x)
\end{aligned}$$

When there is a (fractional) blocking set B, then the definitions are the same except in $P_{\partial A}(x)$ and $\mathsf{Q}(\vec{e} = \{y, x\})$ the $P_A(x)$ is replaced by $P_A(x) + P_B(x)$. Notice all these definitions correspond to the regular definitions when A and B have no partial terms (i.e. $\forall x : P_A(x),\, P_B(x) \in \{0, 1\}$).

We define a *fractional striped set* A by giving $\alpha_0^A,\ \alpha_1^A, \ldots, \alpha_n^A \in [0, 1]$; this defines a $P_A(\cdot)$ by

$$P_A(x) = \alpha_k^A \qquad if\ |x| = \sum_{i=1}^{n} x_i = k\ .$$

A *fractional hamming ball* [BL91c] $A_{k,\alpha}$ is given by $\alpha_{i<k} = 1$, $\alpha_k = \alpha$, and $\alpha_{i>k} = 0$.

With these definitions we are ready to show a modified BCF for the binary hypercube. We will work with exterior blocking conductance and will consider the set B_i as a *blocking set*, by which we mean that defining $C := A^c \setminus B$ then the minimal $Q(A, C)$ occurs when $Q(A, B_i)$ is maximized, i.e. when B_i blocks the most flow from A.

Theorem 6.4. *The BCF for the binary hypercube reduces to the case of fractional hamming balls $A_{k,\alpha}$, with elements of the blocking set only at k or else also at $k+1$ if $\alpha + \alpha_k^B = 1$.*

Proof. By the Corollary, A and B can be assumed to be fractional striped sets, where $\alpha_{i\in\mathcal{I}'}^A = 1$, $\alpha_{i\in\mathcal{I}''}^B = 1$, and both are 0 elsewhere.

First, assume $\lfloor \frac{n}{2} \rfloor \notin A$ and ignore B. We show that fractional hamming balls have minimal flow $Q(A, A^c)$.

We now compress to a fractional hamming ball. Let $M = \max\{i < n/2 : \alpha_i^A > 0\}$ and $m = \max\{i < M : \alpha_i^A = 0\}$. Then let $\alpha_m^A \leftarrow 1$ and $\alpha_M^A \leftarrow \alpha_M^A - \binom{n}{m}/\binom{n}{M}$, i.e. shift units from M to m. Then $\Delta_{\mathsf{Q}(A,C)} \le \binom{n}{m}\left[\left(\frac{m}{n} - \frac{n-m}{n}\right) + \left(\frac{M}{n} - \frac{n-M}{n}\right)\right] < 0$. If $\alpha_M^A < \binom{n}{m}/\binom{n}{M}$ then simply shift units of $M-1$ as well, and the same sort of computation shows $\Delta_{\mathsf{Q}(A,C)} < 0$. Repeat this process until no more compression is possible, i.e. only a fractional hamming ball is remaining. Likewise, compress the $i > n/2$ upward to a second fractional hamming ball.

Suppose the fractional hamming balls are $A_{k,\alpha}$ and $A_{k',\alpha'}$ (the second is centered at $(1, 1, \ldots, 1)_n$). To combine the hamming balls, without loss assume $\pi(A_{k,\alpha}) \ge \pi(A_{k',\alpha'})$, i.e. $k > k'$, or $k = k'$ and $\alpha \ge \alpha'$. Then, as in the previous paragraph, increase α_k^A while decreasing $\alpha_{n-k'}^A$ and as before $\Delta_{\mathsf{Q}(A,C)} \le 0$. Repeat until compressed into a single fractional hamming ball.

Now, assume $\lfloor \frac{n}{2} \rfloor \in A$. Then there is a component in the center, let $m = \max\{i < n/2 : \alpha_i = 0\}$ and $M = \min\{i > n/2 : \alpha_i = 0\}$. Without loss assume $|n/2 - M| \le |n/2 - m|$. Then, as above increase α_m while decreasing α_M and again $\Delta_{\mathsf{Q}(A,C)} \le 0$. Repeat until $\lfloor \frac{n}{2} \rfloor \notin A$, then compress as above.

Consider blocking sets and the stage of compressing into two fractional hamming balls. If $\alpha_{i_{t+1}}^B > 0$ then let $\alpha_M^B = \Delta\alpha_M^A$ and have $\alpha_{i_{t+1}}^B$ decrease accordingly, i.e. replace compressed units of A with units of B. Then

$$\Delta_{\mathsf{Q}(A,C)} \le \binom{n}{m}\left[\left(\frac{m}{n} - \frac{n-m}{n}\right) + \left(\frac{M}{n} - \frac{n-M}{n}\right) + \left(\frac{n-M}{n} - \frac{M}{n}\right)\right] \le 0\ .$$

Likewise, near m shift units from α_m^B to α_{m-1}^B and the calculations work out similarly. If $\alpha_{m-1}^B = 1$ or $\alpha_{m-1}^A = 1$ then leftover blocking units from α_m^B can be distributed anywhere as they are not needed, likewise at α_{M+1}^B.

The blocking set B in other stages of the compression work out similarly.

The problem is now reduced to a fractional hamming ball $A_{k,\alpha}$ with α_k^B and α_{k+1}^B possibly non-zero. Recall in the Lemma that A was increased in size and B decreased in order to round to the nearest level sets S_i. Decreasing $A_{k,\alpha}$ to $\pi(A_{k,\alpha}) = x$ and increasing B, to restore both to their original sizes, can only decrease $\mathsf{Q}(A, C)$. □

Theorem 6.5. *The lazy random walk on the hypercube* $\{0,1\}^n$ *satisfies*

$$\phi(x) \geq \frac{1}{4\,n}\,\log(1/x(1-x))\ .$$

Proof. From Theorem 6.4 we can assume A is a fractional hamming ball $A_{k,\alpha}$.

Let $\lambda = 2^{-n}\,\frac{1}{2}\,\left((1-\alpha)\binom{n}{k-1} + \alpha\binom{n}{k}\right)$. Then $\frac{1}{2}\,(1-\alpha)\binom{n}{k-1}$ term can be optimistically thought of as blocking half the edges from S_{k-1} to S_k, and similarly for $\binom{n}{k}$ and S_k to S_{k+1}. Then, the set B with $\pi(B) \leq \lambda$ blocks at most half of $\mathsf{Q}(A, A^c)$. Also, $\lambda \geq \frac{1}{4}\,\pi(\partial_{int}A)$, because $\pi(\partial_{int}A) \leq \pi(S_{k-1}) + \alpha\pi(S_k)$ and the 1/4 easily follows when $\alpha \leq 1/2$ or $\alpha > 1/2$.

Then $\lambda\,\mathsf{Q}(A, C) \geq \frac{1}{8}\,\pi(\partial_{int}A)\,\mathsf{Q}(A, A^c)$ and Talagrand's inequality [Tal93] $\pi(\partial_{int}A)\,\mathsf{Q}(A, A^c) \geq \frac{1}{4}\,x^2(1-x)^2\,\frac{\log(1/x(1-x))}{n}$ gives the result. □

We could, of course, have found a bound for Theorem 6.5 directly without reference to [Tal93], however the form of the result in [Tal93] is ideal for the following corollary.

Corollary 6.4. *The mixing time of the hypercube* $\{0,1\}^n$ *is*

$$\tau = O(n\,\log n)$$

Chapter 7

Open Problems

Several open problems and conjectures are raised in this thesis.

Problem 7.1. In Chapter 4 the mixing time bounds for the Markov chain on the grid $[k]^n$ and the one on linear extensions – using geometry with transition probability p to bound mixing – gave results $\Theta(p\log(1/p))$ from the correct bounds. Perhaps Theorem 4.9 can be strengthened to

$$\tau = O\left(\frac{p\log(1/p)}{\Phi_g^2}\right) \qquad \text{and} \qquad \rho = \Omega\left(\Phi_g^2/p\right) .$$

One reason to believe this holds is that geometric Markov chains have a lot of the symmetry of product Markov chains. In a d term product Markov chain the mixing time increases by $O(d\log d)$, but in the geometric problems $1/\Phi^2$ overstates this and increases the mixing time by $O(d^2)$. In these problems, $p = \Omega(1/d)$ so $p\log(1/p)$ would correct this overstatement.

Problem 7.2. As mentioned in Chapter 5, Jerrum and Son [JS02] recently studied balanced matroids by using an inductive argument on the second eigenvalue, and similarly on the log-Sobolev constant, to show that

$$\begin{aligned} \lambda &\geq 1/m\,n \qquad \textit{and} \qquad \tau = O(m\,n^2\,\log m) \\ \rho &\geq 1/m\,n \qquad \textit{and} \qquad \tau = O(m\,n\,(\log n + \log\log m)) \ . \end{aligned}$$

These are of the correct orders, as discussed at the end of Chapter 5. Our question is, is it possible to prove this mixing time bound using blocking conductance? Problem 7.3

below suggests a variation on blocking conductance that we expect can be used to obtain the correct mixing time bounds for this problem.

Problem 7.3. In Chapter 6, in both the hypercube example and the direct product example, only the $\log^2(1/\mathsf{P}_{min})$ term kept the mixing time bounds from being correct. A method to bound the mixing time directly in terms of $h_2^+(x)$ might give mixing in

$$\tau = O\left(\int_{\pi_0}^{\frac{1}{2}+\pi_{max}} \frac{dx}{x\, h_2^+(x)^2}\right) \tag{7.1}$$

where $\pi_0 = \min \pi(v)$ and $\pi_{max} = \max \pi(v)$. This would give correct mixing time bounds on a variety of Markov chains such as the line, hypercube, barbell, and products of Markov chains. It would also show that the quantity $\tilde{\beta}_p^2(K)$ gives mixing time bounds as strong as the log-Sobolev constant ρ.

This may also solve the geometry and matroid problems. The difference in the definitions of $\Phi(x)$ and $h_2^+(x)$ is in the transition probability P versus $\sqrt{\mathsf{P}}$. In both the geometric and balanced matroid problems (Chapters 4 and 5), it was found that $\Phi(x) \propto p\,\log(1/x)$. This says if $A \subset K$ and $\pi(A) = x$ then the average $\mathsf{P}(\alpha, A^c)$ for $\alpha \in A$ is proportional to $p\,\log(1/x)$. If the distribution is sufficiently concentrated around this average then it can be expected that $\Phi(x) \propto p\,\log(1/x)$ will become $h_2^+(x) \propto \sqrt{p\,\log(1/x)}$. This would improve the mixing times by $O((p\,\log\log(1/\pi_0))^{-1})$, and for both the geometric and matroid problems it would give optimal mixing time results.

Regardless of whether (7.1) is correct, Theorem 6.3 shows that $h_2^+(x)^2$ and $\phi(x)$ are nearly of the same order, so the heuristics in the previous paragraph suggest that blocking conductance $\phi(x)$ may in fact be sufficient for proving optimal bounds on all these examples.

Problem 7.4. It would be nice if the spectral gap λ could be bounded by blocking conductance. One possible form is

$$\lambda = \Omega\left(\left[\int_{\pi_0}^{1/2+\pi_{max}} \frac{dx}{\phi(x)}\right]^{-1}\right).$$

A heuristic argument for this form is that removing the $1/x$ term in the bound on τ should roughly correspond to removing the $\log(1/\pi_0)$ in the upper bound $\tau \leq \lambda^{-1}\,\log(1/\pi_0)$, and hence the upper bound on τ may become an upper (or bound) on λ.

This form could be used to show some classical results. For example, letting $\phi(x) = \frac{1}{4}\,\Phi^2$ would give $\lambda = \Omega(\Phi^2)$. On the other extreme, if an upper bound on λ were of the same form then when $\phi(x)$ hits the upper bound $\Psi_{ext}(A) \leq 4\,\Phi(A)$ in Theorem 6.3 for all A, then we can let $\phi(x) = \Phi$ and obtain $\lambda = O(\Phi)$. It seems too much to hope that λ be bounded both above and below in this way, but these results do show the possibility that at least one of the bounds may be correct.

Bibliography

[AD86] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93:333–348, 1986.

[AF] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs (book to appear). *URL for draft http://www.stat.Berkeley.edu/users/aldous.*

[Ald83] D. Aldous. Random walks on finite groups and rapidly mixing markov chains. *Séminaire de Probabilités XVII 1981/82*, pages 243–297, 1983.

[ALW97] D. Aldous, L. Lovász, and P. Winkler. Mixing times for uniformly ergodic markov chains. *Stochastic Processes and their Applications*, 71:165–185, 1997.

[BD97] R. Bubley and M. Dyer. Faster random generation of linear extensions. Technical Report 97.41, School of Computer Studies, University of Leeds, August 1997.

[Bez99] S.L. Bezrukov. Edge isoperimetric problems on graphs. In *Graph Theory and Combinatorial Biology*, volume 7, pages 157–197. Bolyai Soc. Math. Stud., 1999.

[BG96] S. Bobkov and F. G otze. Discrete isoperimetric and poincaré-type inequalities. 1996.

[BHT00] S. Bobkov, C. Houdré, and P. Tetali. λ_∞, vertex isoperimetry and concentration. *Combinatorica*, 20(2):153–172, 2000.

[BL91a] B. Bollobás and I. Leader. Compression and isoperimetric inequalities. *Journal of Combinatorial Theory*, 56:47–62, 1991.

[BL91b] B. Bollobás and I. Leader. Edge-isoperimetric inequalities in the grid. *Combinatorica*, 11(4):299–314, 1991.

[BL91c] B. Bollobás and I. Leader. Isoperimetric inequalities and fractional set systems. *Journal of Combinatorial Theory*, 56:63–74, 1991.

[Bob00] S.G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Annals of Probability*, 27(4):1903–1921, 2000.

[Che00] F. Chen. Ph.d. thesis. *Department of Mathematics, Yale University*, 2000.

[DF91] M. Dyer and A. Frieze. Computing the volume of convex bodies : A case where randomness provably helps. *Symposium on Applied Mathematics*, 44:123–169, 1991.

[DFK91] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal Association Computing Machines*, 38(1):1–17, 1991.

[DS91] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1:36–61, 1991.

[DSC93] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

[DSC96] P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[ea] M. Jerrum et. al. Markov chains. *Preprint*.

[FM92] T. Feder and M. Mihail. Balanced matroids. *STOC '92*, pages 26–38, 1992.

[Hel57] I. Heller. On linear systems with integral valued solutions. *Pacific Journal of Mathematics*, 7:1351–1364, 1957.

[Hou01] C. Houdré. Mixed and isoperimetric estimates on the log-sobolev constant of graphs and markov chains. *Combinatorica*, 2001.

[HT96] C. Houdré and P. Tetali. Isoperimetric invariants for product markov chains and graph products. *Preprint*, 1996.

[Jer98] M. Jerrum. Mathematical foundations of the markov chain monte carlo method. In *Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithms and Combinatorics*, volume 16, pages 116–165. 1998.

[Jer00] M. Jerrum. Matroids. *Preprint*, 2000.

[JS88] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of the permanent resolved. *Proc. 20nd Annual ACM Symposium on Theory of Computing*, pages 235–243, 1988.

[JS02] M. Jerrum and J-B. Son. Spectral gap and log-sobolev constant for balanced matroids. *Preprint*, 2002.

[JVV86] M. Jerrum, L. Valiant, and V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretic Computer Science*, 43:169–188, 1986.

[Kan] R. Kannan. Optimization using a random walk. *Preprint*.

[KK91] A. Karzanov and L. Khachiyan. On the conductance of order markov chains. *Order*, 8(1):7–15, 1991.

[KLM02] R. Kannan, L. Lovász, and R. Montenegro. A logarithmic cheeger inequality and mixing in random walks. *Preprint*, 2002.

[KLS99] R. Kannan, L. Lovász, and M. Simonovitz. Isoperimetric problems for convex bodies and a localization lemma. *J. Discr. Comput. Geom.*, 13:282–287, 1999.

[KR01] V.S. Anil Kumar, , and H. Ramesh. Coupling vs. conductance for the jerrum-sinclair chain. *Random Structures and Algorithms*, 18(1):1–17, 2001.

[LK99] L. Lovász and R. Kannan. Faster mixing via average conductance. *STOC '99*, pages 282–287, 1999.

[LS90] L. Lovász and M. Simonovits. Mixing rate of markov chains, an isoperimetric inequality, and computing the volume. *Proc. 31st Annual Symp. on Found. of Computer Science*, pages 346–355, 1990.

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

[LW96] L. Lovász and P. Winkler. Reversal of markov chains and the forget time. Technical Report 1099, Department of Computer Science, Yale University, 1996.

[Mic99] L. Miclo. Relations entre isopérimétrie et trou spectral pour les chaînes de markov finies. *Probability Theory and Related Fields*, 114:431–485, 1999.

[MRR+53] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[MS92] M. Mihail and M. Sudan. Connectivity properties of matroids. Technical Report CSD-92-662, University of California at Berkeley, 1992.

[MS01] R. Montenegro and J-B. Son. Edge isoperimetry and rapid mixing on matroids and geometric markov chains. *STOC*, 2001.

[Mur01] S. Murali. Curvature, isoperimetry, and discrete spin systems. *Ph.D. Thesis*, 2001.

[Sin92] A. Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1(4):351–370, 1992.

[Tal93] M. Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric and Functional Analysis*, 3(3):295–314, 1993.

[Vig99] E. Vigoda. Sampling from gibbs distributions. *Ph.D. Thesis, University of California at Berkeley*, 1999.

[Wil97] D. Wilson. Mixing times of lozenge tiling and card shuffling markov chains. *Preprint*, 1997.