Random walks on finite convex sets of lattice points

Senior thesis

Bálint Virág
Harvard University

April 1, 1996

Introduction

Sometimes parents let their youngest kid roam in the Museum of Very Modern Art, while they are looking at the exhibit "Cubism and Mathematics". The kid wanders around happily and randomly in the other exhibit halls. When it is time to go the parents have to find their child. If she left not so long ago, they believe that she must be somewhere near. If it has been a long time since she left, they think she could be in any exhibit hall, with about equal probability. We are interested in the question: how long is that 'long time'?

It is possible to model the rooms of the Museum of Very Modern Art as a subset of V of the d-dimensional lattice $\mathbb{Z}^d$. The structure of the museum is given by a graph structure on V, which follows naturally from the set of vertices V: two rooms/vertices are connected if their Euclidean distance is 1 (see Figure 1). We assume further that the shape of the accessible exhibit space in the museum is close to convex. The kid's wandering can be very roughly modeled as follows: in each step, she picks one of the $2d$ directions at random, and if there is another room in that direction, she goes there. We are interested in how long it takes for such a random walk to get close to its stationary distribution.

The main theorem in this paper provides a solution to an open problem stated in [1], claiming that $c_d\gamma^2$ steps are sufficient, where γ is the graph-theoretic diameter of V, and c_d is a constant, which depends only on the dimension d. The result was proved by Diaconis and Saloff-Coste [1], for $d = 2$. The proof of this result uses geometric arguments and the techniques developed in [1]. Apart from proving the main theorem, this paper attempts to acquaint the reader with Nash inequalities, which

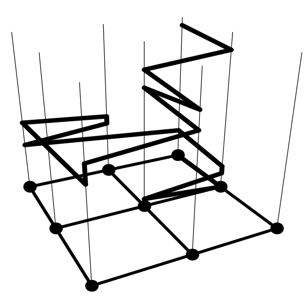

Figure 1: A random walk on $\{1, 2, 3\}^2$. Time is displayed vertically.

we will use to prove our results. We will state and explain the theorems we use without reproducing the original proofs.

The first chapter of this paper introduces a new notion of convexity for sets V of lattice points. Then follows a proof that convex sets contain geodesic paths which are close to straight in the Euclidean sense. Chapter 2 introduces the reader to two techniques used for bounding the convergence of a random walk: the classical second-largest eigenvalue bound and Nash inequalities. Chapter 3 reviews geometric techniques used to estimate the second-largest eigenvalue and to get Nash inequalities. Chapter 4 contains the second part of our original contribution, where we prove that for the random walk defined above, order γ^2 steps are sufficient to reach stationarity.

Many thanks to my advisor, Prof. Persi Diaconis for his support.

Chapter 1. Convexity

1 The definition

This section reviews traditional definitions of convexity on a subset of a lattice, and introduces a new definition, which has most of the properties associated with the notion of convexity in $\mathbb{R}^d$.

As sets of lattice points are naturally equipped with a graph structure, one could define convexity borrowing the definition form graph theory. A convex subgraph G' of a graph G must contain all shortest paths between its vertices. Thus a set $V \subset \mathbb{Z}^d$ can be defined convex if its natural graph is a convex subgraph of $\mathbb{Z}^d$. Unfortunately, under this definition the only convex sets of lattice points are boxes, so it is too restrictive.

One can try to define V to be convex if it is the set of lattice points inside a convex set in $\mathbb{R}^d$. Based on the notion of convexity in $\mathbb{R}^d$, we would like a convex set $V \subset \mathbb{Z}^d$: (1) to be connected in the natural graph structure, (2) not to have any 'bottlenecks', and equivalently (3) to contain a path between any two of its points which is close to 'straight' in the Euclidean sense. This definition violates the first condition even when $d = 2$. There are 'convex' sets, such as $V = \{(z, z) | 0 \le z \le n\}$, whose natural graph is not connected.

To correct this, we can define convex sets $V \subset \mathbb{Z}^d$ to be sets of lattice points inside a convex set in $\mathbb{R}^d$, if they define a connected graph. This way we get a definition that works well in the case $d = 2$. Diaconis and Saloff-Coste use this definition in [1] to prove a two dimensional version of the main theorem of this paper. However, the same definition will not work well in 3 dimensions. A counterexample named 'dog' is given in [1]: Consider the set $Dog_n = \{(x, y, 0) | 0 \le x, y \le n\} \cup \{(x, y, 1) | -n \le x, y \le 0\}$ (see Figure 2). Note that the convex hull of Dog_n contains no extra lattice points. Such a set will inevitably have a 'bottleneck' (it consists of two disjoint 'squares' connected by one edge). Also, any path between the vertices $(0, n, 0)$ and $(-n, 0, -1)$ must contain the vertex at the origin—therefore it cannot be 'straight' in the Euclidean sense.

The definition of convexity presented next will not allow such counterexamples, yet it is intuitive.

Definition 1.1 *A set of lattice points $V \subset \mathbb{Z}^d$ is convex if there is a nonempty convex set $C \subset \mathbb{R}^d$ such that $z \in V \Leftrightarrow d_\infty(z, C) \le \frac{1}{2}$.*

Here $d_\infty(x, y) = \max(|x_i - y_i|)$. We call the convex set C the **base** for the set of lattice points V. There are two equivalent versions that help understand the definition. Let $\mathcal{D}$ denote the closed d_∞ ball of radius $\frac{1}{2}$, which is a hypercube of volume 1. Then:

- V is the set of lattice points inside $C + \mathcal{D}$, where $+$ denotes element-wise addition.
- One can think that each lattice point z has its own cubicle, $z + \mathcal{D}$. These cubicles are almost disjoint (up to lower dimensional components), and they cover $\mathbb{R}^d$. Then V is the set of lattice points whose cubicles intersect the convex set C (see Figure 3).

The following sections will show that convex sets of lattice points are connected and have geodesic paths that are close to straight in the Euclidean sense. This means that our definition satisfies the intuitive criteria for convexity given above. It also shows that the previous counterexample sets Dog_n and $V = \{(z, z) | 0 \leq z \leq n\}$ are not convex. In Section 3 we show that in most cases the convex set of lattice points gained by letting C be an Euclidean line is an single infinite path.

We conclude this section with the most important notation. By a subset V of $\mathbb{Z}^d$ we mean the set of points together with a **natural** graph theoretic structure: for $x, y \in V$ there is an edge (x, y) if their Euclidean distance is 1. For the sake of simplicity we will use the notation V for both the subset of $\mathbb{Z}^d$ and the graph. We will assume throughout that V is finite.

In this paper distances are d_∞, unless stated otherwise. $D(r)$ will denote the closed d_∞-ball of radius r about the origin, and $\mathcal{D}$ denotes $D(\frac{1}{2})$, which has volume 1:

$$D(r) := \{x \in \mathbb{R}^d | d_\infty(x, O) \leq r\} \qquad \mathcal{D} := D(\tfrac{1}{2}). \tag{1}$$

Addition and scalar multiplication in connection with subsets of $\mathbb{R}^d$ will be understood as element-wise operations. For a lattice point $z \in \mathbb{Z}^d$ we will call the set $z + \mathcal{D} := \{z\} + \mathcal{D}$ the **cubicle** of z. Note that the collection of all cubicles covers the entire $\mathbb{R}^d$. For two points x, y, we will distinguish the line $\overline{xy}$ and the line segment xy. We will also use the fact that if A and B are convex subsets of $\mathbb{R}^d$, then $A + B$ is convex.

2 Alternative definitions

The following lemma provides alternative formulations of our definition. We will need these in later sections.

Lemma 2.1 *For a bounded subset V of the lattice $\mathbb{Z}^d$ the following are equivalent:*

1. *V is convex.*
2. *There is a closed convex set C' in $\mathbb{R}^d$ and an $\varepsilon > 0$ such that: (1) V is the set of lattice points at most $\frac{1}{2}$ away from C'. (2) there are no other lattice points in the $\frac{1}{2} + \varepsilon$ neighborhood of C'*

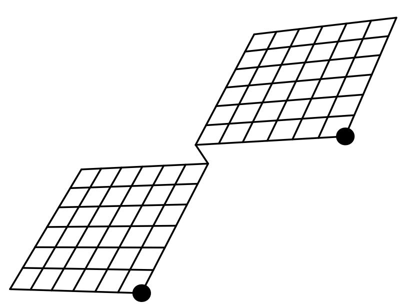

Figure 2: A small dog

3. *There is a closed convex set C'' in $\mathbb{R}^d$ and an $\varepsilon > 0$ such that: (1) V is the set of lattice points at most $\frac{1}{2} - \varepsilon$ away from C''. (2) there are no other lattice points in the $\frac{1}{2}$ neighborhood of C''*

Proof.

($1 \Rightarrow 2$) For each $p \in V$, let c_p be a point in the intersection of $p + \mathcal{D}$ and C (C as in the definition of convexity (1.1)). Let C' be the convex hull of all c_p-s. Let V' be the set of lattice points with d_∞-distance at most $\frac{1}{2}$ from C'. Then, by construction of C' we have $V \subset V'$. Also the convex hull C' of the c_p-s is contained in C, because C is convex, so $V' \subset V$. Since C' is closed and bounded, there is a lattice point outside V with minimal d_∞-distance δ from C'. Let $\varepsilon = (\delta - \frac{1}{2})/2$. This will work.

($2 \Rightarrow 3$) Take C'' to be the set of points with d_∞ distance at most ε from C'.

($3 \Rightarrow 1$) Clear. □

3 Euclidean lines and paths in $\mathbb{Z}^d$

This section shows that for most Euclidean lines C, the convex set of lattice points induced by C as a base is an infinite path. First, we need to introduce some notation. Let $\mathbb{Z} + \frac{1}{2}$ be the set of reals of the form integer$+\frac{1}{2}$. Let E_n be the set of points in $\mathbb{R}^d$ for which at least n coordinates are in $\mathbb{Z} + \frac{1}{2}$. Geometrically, E_n is the union of all $(d-n)$ and lower dimensional components of the cubicles $z + \mathcal{D}$ for $z \in \mathbb{Z}^d$. As an example, in 3 dimensions E_2 is the union of all edges. We call a line $\overline{v}$ a **proper line** if it does not intersect the set E_2. Define the **infinite path** to be the natural graph of $\mathbb{Z}$.

Informally, most lines are proper, and Lemma 4.2 shows that there is a proper line 'close' to any line.

Proposition 3.1 *Let C be a proper line, and let V be the set of lattice points whose cubicles intersect C. Then the natural graph of V and the infinite path are isomorphic as graphs. Moreover, under this isomorphism, any section $(j, j+1, ...k)$ of the infinite path maps to a geodesic path of $\mathbb{Z}^d$.*

The proof of this result requires some lemmas concerning proper lines. Let $\mathcal{D}^o$ denote the interior of the hypercube $\mathcal{D}$.

Lemma 3.2 *A proper line $\overline{v}$ is not contained in any of the hyperplanes $\xi_i = z_h$, where $z_h \in \mathbb{Z} + \frac{1}{2}$.*

Proof. WLOG suppose that $\overline{v}$ is contained in the hyperplane $\xi_1 = \frac{1}{2}$. Then for some $a, b \in \mathbb{R}^d$ we can write our line as the set $\overline{v} = \{a + \lambda b | \lambda \in \mathbb{R}\}$, where $a_1 = \frac{1}{2}$, $b_1 = 0$, and some coordinate b_i is nonzero. Therefore if we take $\lambda = \frac{1}{2} - a_i/b_i$, we get a point on $\overline{v}$ which has two coordinates equal $\frac{1}{2}$, a contradiction. □

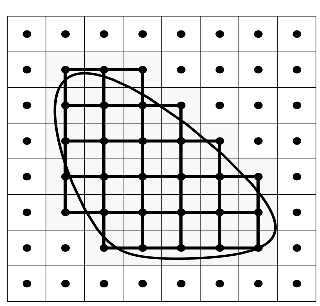

Figure 3: A convex set of lattice points and its natural graph

Lemma 3.3 *Suppose that a proper line $\overline{v}$ intersects a cubicle $z + \mathcal{D}$ in a nonempty set I. Then I is a closed interval in $\overline{v}$, containing at least two points, and at least one point from $z + \mathcal{D}^o$.*

Proof. Let v be a unit vector parallel to $\overline{v}$. Since both $\overline{v}$ and $z + \mathcal{D}$ are closed and convex, their intersection must be a closed and convex subset of $\overline{v}$, and so a closed interval. Note that the cubicle $z + \mathcal{D}$ is determined by $2d$ inequalities of the form $\pm\xi_i \leq \pm z_i + \frac{1}{2}$. Let $p \in I$, then p has at most one coordinate of the form integer$+\frac{1}{2}$. Thus there is an ε-ball about p which satisfies all inequalities for the cubicle except for one. Then one of $p \pm \frac{\varepsilon}{2}v$ satisfies that inequality as well, call this point p'. p' must be in the interior of $z + \mathcal{D}$, since otherwise both p and p' were contained in the same plane $\xi_i = z_i \pm \frac{1}{2}$. By Lemma 3.2 this would contradict the fact that v is a proper line. $\square$

Corollary 3.4 *A proper line $\overline{v}$ intersects a cubicle $z+\mathcal{D}$ if and only if it intersects its interior $z+\mathcal{D}^o$. Also, $\overline{v}$ intersects $z + \mathcal{D}^o$ in an open interval.*

Lemma 3.5 *Let $z \neq w \in \mathbb{Z}^d$, and suppose that the cubicles $z + \mathcal{D}$ and $w + \mathcal{D}$ intersect in a point $p \notin E_2$. Then z and w are connected by an edge in the natural graph of $\mathbb{Z}^d$.*

Proof. Note that the d_∞ distance of z and w is at most 1. Therefore the difference $|z_i - w_i|$ is at most 1, and if the two vertices are not connected, they differ in at least two coordinates i, j. Consider the hyperplanes $\xi_i = (z_i - w_i)/2$ and $\xi_j = (z_j + w_j)/2$. Both separate the hypercubes $w+\mathcal{D}$ and $z+\mathcal{D}$, so the common point p must be contained in both planes. This means that p has two coordinates of the form integer$+\frac{1}{2}$, so $p \in E_2$, a contradiction. $\square$

Proof of Proposition 3.1. Isomorphism part. For any distinct $z, z' \in V$, the sets $z + \mathcal{D}^o$ and $z' + \mathcal{D}^o$ are disjoint, and they intersect C in disjoint nonempty open intervals (Corollary 3.4). For each $z \in V$, pick a point $p(z)$ in $(z + \mathcal{D}^o) \cap C$. Then the points $p(z)$ form a discrete set on the line C. We can define an order on them in one direction along the line C, and label them with integers so that $p_i < p_{i+1}$ for all $i \in \mathbb{Z}$. This defines a similar order on the corresponding z-s, which is independent of the choice of $p(z)$-s. We claim that the map $i \mapsto z_i$ is a graph isomorphism. We have to prove that (z_i, z_j) is an edge if and only if $i = j \pm 1$.

(If) Consider two consecutive points p_i, p_{i+1}, and the corresponding z_i, z_{i+1}. There is no cubicle $z_j + \mathcal{D}$ that intersects the line between p_i and p_{i+1}, since then we would have $i < j < i + 1$ by our ordering. Also, the cubicles $z_i + \mathcal{D}$ cover the line C. Therefore $z_i + \mathcal{D}, z_{i+1} + \mathcal{D}$ cover the line segment p_ip_{i+1}, and since they are closed, they must intersect on the line segment (otherwise they would provide a separation of a connected segment). This means that $z_i + \mathcal{D}$ and $z_{i+1} + \mathcal{D}$ have a common point not in E_2, so by Lemma 3.5 z_i and z_{i+1} are connected by an edge in the natural graph of V.

(Only if) Let (z_i, z_j) be an edge in V, then z_i and z_j differ in one coordinate by 1. Therefore the set $(z_i + \mathcal{D}) \cup (z_j + \mathcal{D})$ is convex, and so it contains the entire line segment p_ip_j. Note that the set $(z_i + \mathcal{D}) \cup (z_j + \mathcal{D})$ is disjoint from the interior of the cubicle of any third lattice point. Therefore no points on the line segment p_ip_j can be in the interior of a third cubicle. In particular, there is no other point p_k on this segment. This means that in our labeling p_i and p_j had to be consecutive points, so $i = j \pm 1$.

Proof that the sections are geodesic. Suppose that a section $z_j, z_{j+1}, ..., z_k$ is not geodesic, then it contains edges (z_a, z_{a+1}) and (z_b, z_{b+1}) for $a < b$, such that for one of the coordinates (WLOG the first one) we have:

$$(z_a)_1 = (z_{b+1})_1, \quad \text{and} \quad (z_{a+1})_1 = (z_b)_1 = (z_a)_1 \pm 1 \quad (\text{WLOG} = (z_a)_1 + 1).$$

This means that the cubicles $z_a + \mathcal{D}^o$ and $z_{b+1} + \mathcal{D}^o$ are strictly separated from $z_b + \mathcal{D}^o$ by the hyperplane $\xi_1 = (z_a)_1 + \frac{1}{2}$. Therefore the line C, which contains the points p_a, p_b, p_{b+1} in this order, must intersect that hyperplane twice (once between p_a and p_b, and once between p_b and p_{b+1}). Thus C is contained in the hyperplane, so it cannot be a proper line, a contradiction. $\square$

Corollary 3.6 *(of the proof.) Let $x, y \in \mathbb{Z}^d$, and x', y' be points in the interior of their cubicles respectively. Assume further that the line $\overline{x'y'}$ is proper. Then there is a geodesic path $\gamma_{x,y} = (x = z_1, z_2, ..., z_j = y)$ such that the cubicles z_i intersect the line segment $x'y'$.*

4 Straight paths

Corollary 3.6 showed that a segment of a proper line defines a geodesic path in $\mathbb{Z}^d$. This section uses this result to prove that there are almost straight geodesic paths in any convex set of lattice points V.

Proposition 4.1 *In a convex set of lattice point V any two vertices x, y are connected by a path $\gamma_{x,y}$ in V with the following properties: (1) $\gamma_{x,y}$ is geodesic as a path in $\mathbb{Z}^d$ (2) each vertex in $\gamma_{x,y}$ has d_∞ distance less than 1 from the Euclidean line $\overline{xy}$.*

Given two lattice points $x, y \in V$, our strategy is to find a nearby x' and y' in the base C of V such that the segment $x'y'$ is proper, and so the path that it defines as a base will work. We cannot do this directly, because an arbitrary base for V does not always contain a proper line segment. The proof therefore requires a special base for V provided by Lemma 2.1. The following lemma will also be needed:

Lemma 4.2 *Let x, y be points in $\mathbb{R}^d$, and let $\varepsilon > 0$. Then there are points x', y' in the ε-neighborhood of x, y respectively, such that the line $\overline{x'y'}$ is proper.*

Proof. Pick $x' \in x + D(\varepsilon)$ to have rational coordinates which are not in $\mathbb{Z} + \frac{1}{2}$. Then pick z in the ε-neighborhood of $y - x'$ such that the ratios z_i/z_j are all irrational. This means that there are no points with two rational coordinates on the line $\overline{Oz}$. If we set $y' = z + x'$, then there will be no points with two rational coordinates except for x' on the line $\overline{x'y'}$. So the line $\overline{x'y'}$ can only intersect E_2 in x'. But $x' \notin E_2$, so $x'y'$ is a proper line. $\square$

Proof of Proposition 4.1. By Lemma 2.1 there is a convex set C and an $\varepsilon > 0$ such that V is the set of lattice points inside $C + \mathcal{D}$ and there are no extra lattice points inside $C + D(\frac{1}{2} + \varepsilon)$ (recall the definitions of $D(r)$ and $\mathcal{D}$ (1). This means that C and $C + D(\frac{1}{2} + \varepsilon)$ are both base sets for V.

Let x, y be lattice points in V. Then x is inside $C + \mathcal{D} = C + D(\frac{\varepsilon}{2}) + D(\frac{1}{2} - \frac{\varepsilon}{2})$. Therefore there is an $x^0 \in C + D(\frac{\varepsilon}{2})$ such that $d_\infty(x, x^0) \leq \frac{1}{2} - \frac{\varepsilon}{2}$. We can pick y^0 in a similar fashion. By Lemma 4.2 there are points x', y' in the $\frac{\varepsilon}{2}$ neighborhood of x^0 and y^0 such that the line $\overline{x'y'}$ is proper.

We have picked x' and y' so that (1) the points x', y', are contained in the base $C + D(\varepsilon)$ for V, and (2) $d_\infty(x, x') < 1$, so x' is contained in the interior of the cubicle $x + \mathcal{D}$, and the same holds for the y-s. By convexity, $C + D(\varepsilon)$ contains the line segment $x'y'$. This means that Corollary 3.6 applies, and there is a geodesic path $\gamma_{x,y}$ in $\mathbb{Z}^d$, so that for each $z \in \gamma_{x,y}$ the cubicle $z + \mathcal{D}^o$ intersects the segment $x'y'$ and therefore the base $C + D(\varepsilon)$. Thus $\gamma_{x,y}$ is a path in V.

The endpoints of the segments xy and $x'y'$ have d_∞-distance less than $\frac{1}{2}$. It follows that any point in $x'y'$ is less than $\frac{1}{2}$ away from the line $\overline{xy}$, and so for any vertex $z \in \gamma_{x,y}$, $d_\infty(z, \overline{xy}) < 1$. $\square$

Chapter 2. Convergence bounds

The purpose of this chapter is to introduce the reader to the techniques used for bounding the convergence of a Markov chain to its stationary distribution. A proposition will show that the convergence of a random walk is essentially exponential. Later we will argue that it is still worthwhile to try and give polynomial bounds of the form cn^{-d}. We begin with a brief review of symmetric random walks.

5 Random walks on graphs

Informally, a random walk on a graph G is a process in which one starts on a vertex v_1, chooses one of the edges containing v_1 with probability proportional to the weights on the edges, proceeds to the other vertex v_2 along the edge, and repeats this process for v_2, then v_3, etc. All the information about the underlying graph can be described in a $k \times k$ matrix K, where the x, y entry is the probability of going to state y if we are in state x. Note that if s is a probability distribution on the vertex set of G, then the distribution after one step from s is just sK. For any random walk K there is a stationary distribution π for which $\pi K = K$, that is the distribution π does not change after one step.

Random walks (or Markov chains, these terms are interchangeable) can also be given by a matrix *(Markov kernel)* K with positive entries and row sums of 1. Any such Markov chain is a random walk on a directed graph with weighted edges. We call a random walk a *symmetric* or *reversible* if it is a walk on a non-directed graph.

For a symmetric random walk, a stationary distribution can be given by the expression $\pi_x = d(x)/2E$, where $d(x)$ is the sum of the weights on the edges containing the vertex x, and E is the sum of all weights on all edges. In particular, for an unweighted graph (where all weights are 1), $d(x)$ is the degree of x, and E is the number of edges. Note that for a random walk on an undirected graph:

$$\pi_x K(x, y) = \pi_y K(y, x) \qquad \text{for all vertices } x, y. \tag{2}$$

This expression provides an equivalent definition of reversibility, that is, a random walk is reversible if (2) holds for all stationary distributions π. If a chain is reversible and irreducible (the underlying graph is connected), there is a unique stationary distribution. If the chain is reversible, irreducible and aperiodic (the underlying graph is not bipartite), then sK^n converges to the stationary distribution π for any distribution s. As an example, the random walk described in the introduction has the Markov kernel:

$$K(x, y) = \begin{cases} 1/2d & \text{for } x \neq y \text{ neighboring points} \\ 1 - n(x)/2d & \text{for } x = y \end{cases} \tag{3}$$

where $n(x)$ is the number of neighbors of x in the graph V. The Markov kernel K describes the natural random walk on the unweighted graph of V with self-loops attached at the boundary (see Figure 4). These self-loops ensure that every vertex in the underlying graph of K has the same degree. Therefore this walk has *uniform* stationary distribution $\pi_x \equiv 1/k$, where k is the number of vertices in V. The walk K is reversible, and the expression (2) reduces to $K(x, y) = K(y, x)$.

In this paper all Markov chains are assumed to be reversible, irreducible and aperiodic. Let K be the Markov kernel of such a chain. Note that the distribution after n steps from an initial distribution s is given by sK^n. Our goal here is to give an upper bound on the distance of sK^n from the stationary distribution π for a given n. Unfortunately, there are several ways to define distance between these two distributions, all of them having their own right to exist. We introduce three distances here, explaining their particular relevance.

For two distributions, s, s' on V it is conventional to consider the **total variation distance** (TVD) defined by:

$$||s-t||_{TV} = \max_{A \subset V} |s(A) - t(A)| = \frac{1}{2} \sum_{x \in V} |s_x - t_x|.$$

The first expression refers to the measure-theoretic foundations of TVD. For our purposes, we can just think of TVD as an ℓ_1-distance. This is a widely accepted way of defining distance between two distributions, one of its advantages being that it has no privileged distribution. Bounds for the TVD are usually given by bounds on the ℓ_2-distance defined by the inner product:

$$< f, g >= \sum_{x \in V} f(x) g(x) \pi_x \tag{4}$$

Note that this inner product uses the stationary distribution π as a measure. Using Cauchy-Schwarz, one can show that the implied ℓ_2-norm provides a bound on the total variation distance as:

$$2||s-t||_{TV} \leq ||s/\pi - t/\pi||_2. \tag{5}$$

This norm proves to be very useful for graph theoretic methods, when we need to consider a distribution as a measure on functions $V \to \mathbb{R}$. Another scalar product can be defined as:

$$\prec s, t \succ := \sum_{x \in V} s_x t_x \pi_x^{-1}$$

which defines a norm related to the ℓ_2 norm:

$$\wr\wr s - t \wr\wr = ||s/\pi - t/\pi||_2. \tag{6}$$

This norm is useful for the analysis of the geometric properties of the space of probability distributions, and the action of the Markov kernel K on that space. The following sections contain a more detailed analysis. Applying the comparisons 5, 6 to the convergence of random walks we get:

$$2||sK^n - \pi||_{TV} \leq ||(sK^n/\pi) - 1||_2 = \wr\wr sK^n - \pi \wr\wr.$$

We are interested in the maximum of the above distances over the set of distributions S. For a given Markov chain, we introduce the convergence series:

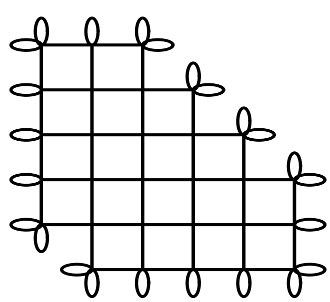

Figure 4: Loops on the boundary

$$D_n = \max_{s \in S} ||sK^n - \pi||^2$$

Let σ_x be the distribution which assigns mass 1 to the vertex x. Then any possible distribution can be written as a linear combination $\sum s_i \sigma_i$ for $0 \leq s_i \leq 1$ and $\sum s_i = 1$. This means that as a subspace of $\mathbb{R}^k$, S is a k-dimensional simplex, the convex hull of the set of distributions σ_x. We will use the abbreviation $K_x^n := \sigma_x K^n$. This section concludes with a proof of a simple fact, which provides an alternative expression of D_n.

Lemma 5.1 *Under any norm on $\mathbb{R}^k$, the worst convergence can be achieved starting from one of the extremal distributions σ_x. That is, for* any *norm '$||$' we have:*

$$\max_{s \in S} ||sK^n - \pi|| = \max_{x \in V} ||K_x^n - \pi||$$

Proof. Let $s = \sum s_x \sigma_x$, then we can write $s - \pi = \sum s_x(\sigma_x - \pi)$. Then $sK^n - \pi = (s - \pi)K^n = \sum s_x[(\sigma_x - \pi)K^n]$. The norm of this is bounded above by $s_x ||\sigma_x - \pi K^n||$, and since the sum of the s_x is 1, this is not more than the expression on the right hand side. Thus LHS$\leq$RHS, and the other direction is clear. □

Thus we can write our convergence series as:

$$D_n = \max_{x \in V} ||(K_x^n/\pi) - 1||_2^2 = \max_{x \in V} \wr K^n - \pi \wr^2.$$

6 Eigenvalue bounds

In the previous section we have seen that the Markov kernel K, as a left linear operator, maps the simplex of distributions S into itself. In this section we will rely heavily on the geometric properties of this simplex. In the previous section we defined a scalar product on $\mathbb{R}^k$ by:

$$\prec s, t \succ := \sum_{x \in V} s_x t_x \pi_x^{-1}.$$

Since the π_x-s are positive, this defines the norm $\wr s \wr := \prec s, s \succ^{\frac{1}{2}}$. Moreover, for any vector $s \in S$, note that

$$\prec \pi, s \succ = \sum_{x \in V} s_x \pi_x \pi_x^{-1} = 1.$$

In particular, $\wr \pi \wr = 1$. This implies that $\prec \pi, s - \pi \succ = 0$, so the decomposition $s = \pi + (s - \pi)$ is orthogonal. Thus the subset $S_0 = \{s - \pi | s \in S\}$ spans a *(k-1)*-dimensional subspace of $\mathbb{R}^k$, and together with the perpendicular π they span the entire $\mathbb{R}^k$. Also, K maps S_0 into itself: $(s - \pi)K = sK - \pi$, where $sK \in S$, since K maps S into itself.

Note that due to reversibility, K is self-adjoint:

$$\prec sK, t \succ = \sum_{x,y \in V} s_x K(x, y) t_y \pi_y^{-1} = \sum_{x,y \in V} s_x \pi_y K(y, x) \pi_x^{-1} t_y \pi_x = \prec s, tK \succ$$

.

Therefore K has an orthonormal basis of eigenvectors $\pi = v_0, v_1, ..., v_{k-1}$. π is perpendicular to all other eigenvectors, so they are contained in the subspace spanned by S_0. If $\pi_x > 0$ for all x, (which holds for all irreducible random walks, and certainly holds for the uniform stationary distribution), then the point π is contained in the k-dimensional interior of the simplex S. This means that in the subspace spanned by S_0, S_0 contains a ball about the origin. Now if we take c small such that cv_i is contained in this ball, then $cv_i + \pi$ will be a probability distribution. Since $(cv_i + \pi)K^n$ is bounded, the absolute value of the eigenvalue β_i is bounded by 1. Thus K has eigenvalues $1 = \beta_0 \geq \beta_1 \geq ... \geq \beta_{k-1} \geq -1$. We define $\beta_* := \max(|\beta_0|, |\beta_{k-1}|)$. Informally, the least eigenvalue β_{k-1} is responsible for periodicity, and for most Markov chains, $\beta_* = \beta_1$. We can write any $s \in S$ in terms of the basis of eigenvectors as:

$$s = \sum_{i=0}^{k-1} c_i v_i, \qquad \text{and so} \qquad sK^n = \sum_{i=0}^{k-1} c_i v_i K^n = c_i \beta_i^n v_i, \tag{7}$$

where $c_i = \prec s, v_i \succ$, and $c_0 = 1$. So the distance of sK^n from the stationary distribution satisfies:

$$\wr\, sK^n - \pi \,\wr^2 = \wr \sum_{i=0}^{k-1} c_i \beta_i^n v_i - \pi \,\wr^2 = \wr \sum_{i=1}^{k-1} c_i \beta_i^n v_i \,\wr^2 = \sum_{i=1}^{k-1} c_i^2 \beta_i^{2n} \tag{8}$$

Using $\beta_i^2 \leq \beta_*^2$ gives us the bound:

$$\wr\, sK^n - \pi \,\wr^2 \leq \sum_{i=1}^{k-1} c_i^2 \beta_*^{2n} = \wr\, s - \pi \,\wr^2 \beta_*^{2n} \tag{9}$$

.

This bound estimates the mapping K^n with a contraction by a factor β_*^n. Note that the distance of any distribution s from π satisfies:

$$\wr\, s - \pi \,\wr^2 = \wr\, s \,\wr^2 + \wr\, \pi \,\wr^2 - 2 \prec s, \pi \succ = \wr\, s \,\wr^2 - 1 \tag{10}$$

Thus if s is one of the extremal distributions σ_x, the factor on the right of (8) can be computed as $\wr\, \sigma_x - \pi \,\wr^2 = \pi_x^{-1} - 1$. This and (5) give us the most widely used form of the eigenvalue bound:

$$||K_x^n/\pi - 1||_2 \leq \sqrt{\pi_x^{-1} - 1}\beta_*^n < \pi_x^{-\frac{1}{2}} \beta_*^n.$$

The following proposition gives a lower bound to estimate the error of the β_*-bound for chains with uniform stationary distribution. It basically says that knowing β_* determines convergence up to a factor of $\log k$.

Proposition 6.1 *For any reversible Markov kernel K with uniform stationary distribution, we have:*

$$\beta_*^n \leq \max_{s \in S} \wr\, sK^n - \pi \,\wr \leq \sqrt{k-1}\beta_*^n$$

Proof. Let $s \in S$ be a distribution, and let us write it in terms of the eigenvectors as in (7). Assume that $\beta_* = |\beta_1|$ (the case when $\beta_* = |\beta_{k-1}|$ is similar). Using the expression (8), we can write

$$\prec s, v_1 \succ^2 \beta_1^{2n} \leq \sum_{i=1}^{n-1} \prec s, v_i \succ^2 \beta_i^{2n} = \wr\, sK^n - \pi \,\wr^2 \leq \wr\, s - \pi \,\wr^2 \beta_1^{2n}$$

The last inequality comes from the β_*-bound (9). Taking square roots and maximizing over s we get

$$\max_{s\in S} | \prec s, v_1 \succ ||\beta_1|^n \leq \max_{s\in S} \wr sK^n - \pi \wr \leq \max_{s\in S} \wr s - \pi \wr |\beta_1|^n.$$

This implies the upper bound part of our theorem, since the maximum of $\wr s - \pi \wr$ is achieved at $s = \sigma_x$ (Lemma 5.1), where it takes the value $\sqrt{k-1}$ (see 10). We want the left inequality to hold for all choices of the unit eigenvector v_1, so we have to take the minimum of the expression on the LHS in terms of v_1. The task is to evaluate the expression:

$$\min_{\substack{\wr v_1 \wr = 1 \\ \prec v_i, \pi \succ = 0}} \max_{s\in S} | \prec v_i, s \succ |.$$

This quantity has a geometric interpretation. Let $\overline{v}$ be a line containing the point π and parallel to the vector v, and let p_v be the projection to the line $\overline{v}$. Note that the absolute value of the scalar product equals the length of the projection: $| \prec v_i, s \succ | = \wr \pi - p_v s \wr$. So we want to evaluate

$$\min_{\substack{\wr v_1 \wr = 1 \\ \prec v_i, \pi \succ = 0}} \max_{s\in S} \wr \pi - p_v s \wr \tag{11}$$

For a geometric interpretation, note that $\max_{s\in S} \wr \pi - p_v s \wr$ equals the distance between π and the farthest point of $p_v S$. We can think that the simplex S casts a shadow on the line $\overline{v}$. The length of the shadow can be defined as the distance of the farthest dark point form the point π. Then the expression (11) is the length of the shortest shadow that the simplex casts as we change the direction of the line $\overline{v}$. Let $S' := 2\pi - S$, that is the reflection of the simplex S about its center of gravity π. Consider the convex hull $H := \text{hull}(S \cup S')$. Just as an example, in 2 dimensions H is a regular hexagon, in 3, it is a cube. Note that

$$p_v H = p_v(\text{hull}(S \cup S')) = \text{hull}(p_v(S \cup S')) = \text{hull}(p_v S) \cup [2\pi - \text{hull}(p_v S)]$$

This means that the expression (11) does not change if we take the maximum over $s \in H$, and not $s \in S$. But the length of the shortest shadow of H is at least as long as the radius of the maximal hypersphere about π that is still contained in H (the two quantities are in fact equal, but won't need this). The following lemma provides the last step in this proof.

Lemma 6.2 *Let S be a d-dimensional regular simplex, and let H be the convex hull of $S \cup -S$, where $-S$ is the reflection of S about its center of gravity. If the radius of the circumsphere of H is 1, the radius of its inscribed sphere is $d^{-\frac{1}{2}}$ for d even and $(d+1)^{-\frac{1}{2}}$ for d odd.*

Proof. We can assume that the center of gravity of S is the origin, and let $a_0, ..., a_d$ be the vertices of S. Then H is a d-dimensional polyhedron, with vertices among $a_0, ..., a_d, -a_0, ... - a_d$. Any d-tuple T of these points determines a plane, which divides the space into two parts. T is a face of the convex polyhedron H if: (1) it does not contain the origin, and (2) all the rest of the vertices are in the same part of the divided space as the origin.

It is a fact from geometry that a point and the origin are on the same side of the plane determined by set of d points T iff when we write p in terms of the basis T, the sum of the coordinates is less than 1. It follows that for any d-tuple T, reflections of the vectors in T are on the same side of T as the origin (the sum of coordinates is -1). This leaves only two other vertices to check.

A face cannot contain both the vertex a and $-a$, since then it would also contain the origin. Also, faces should become faces under permutation of vertices. Up to permutation, the only planes we are left to check are:

$$T_i = (a_1 ... a_i, -a_{i+1}, ..., -a_d) \quad \text{for } i = 0...d.$$

By the above, T_i determines a face iff the sum of the coordinates in the expression of $\pm a_0$ is at most 1. Note, that $\sum_{i=0}^{d} a_i = 0$. This means that when we write a_0 in terms of the vectors in the d-tuple T_i, the sum of coordinates will be $d - 2i$. So we need $d - 2i \leq 1$ and $2i - d \leq 1$, so $i = d/2$ or $i = (d \pm 1)/2$ depending on the parity of d. Note that the cases $i = (d \pm 1)/2$ represent faces that are reflections of each other up to permutation.

The radius of the inscribed circle about the origin is the distance of the closest face. For d even, the vector $v_e = (a_1, ..., a_{d/2}, a_{d/2+1}, ..., -a_d)/d$ is easily checked to be contained in and perpendicular to the plane determined by $T_{d/2}$. For d odd, note that a_0 is contained in $T_{(d-1)/2}$ (the sum of the coordinates is 1). Then the vector $v_o = (a_0, ..., a_{(d-1)/2}, -a_{(d+1)/2}, ..., -a_d)/(d+1)$ is contained in and perpendicular to the plane determined by $T_{(d-1)/2}$. The desired radius is thus the length of the appropriate vector v_o or v_e. Since the length of a_i equals the radius of the circumsphere, this gives the stated result. □

In our case, the radius of the circumsphere is $\wr \sigma_i - \pi \wr = \sqrt{k-1}$. Setting $d = k - 1$ and using the above lemma imply the stated lower bound. □

7 Polynomial decay

The previous section has shown that the distance D_n of a Markov chain from its (uniform) stationary distribution is bounded from above and below by two exponential series, one constant times the other. Although this constant is large, ($\sqrt{k-1}$, where k is the number of states), this still seems to imply that the convergence of the chain is essentially exponential.

The purpose of this section is to convince the reader that for several Markov chains, in the early phase, (that is for all $n < N$ for a given N), D_n is looks more like a polynomial in n^{-1} than an exponential function. This realization provides some motivation to give certain non-exponential bounds on the distance D_n in is early phase. The next section investigates what Nash inequalities have to do with such bounds.

In order to see how the polynomial decay occurs, we will consider a simple random walk on a k-point path $(1...k)$. The transition probabilities are given by:

$$K(x, y) = \begin{cases} 1/2 & \text{if } |x-y| = 1, x = y = 1 \text{ or } x = y = k \\ 0 & \text{otherwise} \end{cases}$$

We will consider the random walk starting at the point 1. Let $s_n = \sigma_1 K^n$ denote the probability distribution after n steps, and $s_{n,i}$ denote the probability of being at point i after n steps. The reader can convince himself looking at Figure 5 that in the early phase, $s_{n,i}$ behaves similarly to the entries of a Pascal triangle, and for $n < k$ we can write

$$s_{n,i} = \frac{1}{2^n} \binom{n}{\lceil \frac{n-i}{2} \rceil}.$$

In particular, for fixed $n < k$, the numbers are $s_{n,i}$ gained by reordering the binomial coefficients $\binom{n}{0}, \binom{n}{1}, ..., \binom{n}{n}$, and dividing them by 2^n (or s_n in this early phase is close to a half bell curve). Using this one can express the distance $\wr s_n - \pi \wr$. By (10), we have

$n\backslash i$	1	2	3	4	5	6 ...
0	1					
1	1	1				
2	2	1	1			
3	3	3	1	1		
4	6	4	4	1	1	
5	10	10	5	5	1	1
$\vdots$						

Figure 5: $2^n s_{n,i}$—that is 2^n times probability distribution on the k-point path after n steps, if the random walks starts at the point 1. The recursion is the similar to that of the binomial coefficients.

$$\wr s_n - \pi \wr^2 = \wr s_n \wr^2 - 1 = \sum_{i=0}^{n} \binom{n}{i}^2 2^{-2n} k^{-1} - 1.$$

Using the identity $\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}$, and the approximation $\binom{2n}{n}/2^{2n} \approx (\pi n)^{-\frac{1}{2}}$ gives the result:

$$\wr s_n - \pi \wr^2 \approx \frac{(\pi n)^{\frac{1}{2}}}{k} - 1,$$

which shows that early the decay of the distribution can be described by an expression of the form $cn^{-\frac{1}{2}} - 1$ (see Fig 6). The early decay is similar for boxes of size n in d dimensions, where we get an approximation of the form $cn^{d/2} - 1$. It is therefore natural to search for a way to prove bounds on s_n of such form for general Markov chains. To conclude this section, we formulate a heuristic image:

Heuristic image 7.1 *Given a random walk on a nice subset of a lattice, there are (nontrivial) constants d, c, N such that for any state s, the norm $\wr sK^n \wr$ is bounded by an expression of the form cn^{-d} for $0 \leq n \leq N$.*

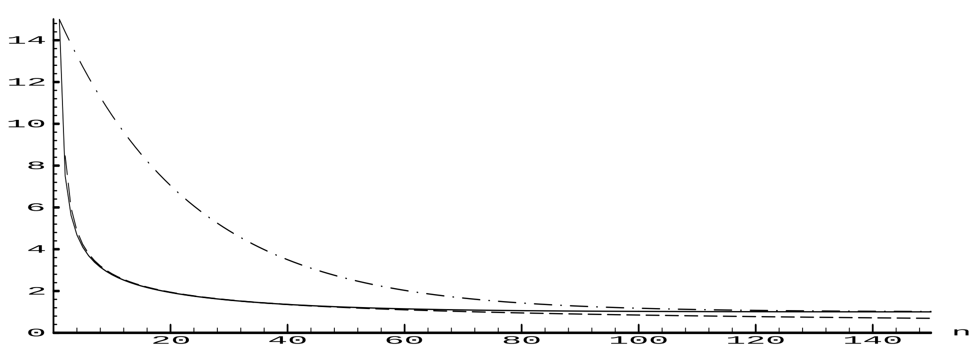

Figure 6: Decay of a distribution for a random walk on a 15-point path. The line (—) stands for $\wr sK^n \wr^2$, (-.-.-) denotes the exponential eigenvalue bound, and (- - -) refers to the polynomial bound, which is better in the early phase.

8 Nash inequalities

Convergence bounds for random walks on graphs are usually proved by looking at the structure of the underlying graph. As we will see later, the structure of the graph is well reflected in the relation between quantities like the norm-square of a distribution $\wr s \wr^2$ and the norm-square of the distribution after one step $\wr sK \wr^2$. Iterating relations of this kind can lead to a bound on the norm $\wr sK^n \wr^2$. Note that by (10) this is just 1 more than the distance-square of sK^n from stationarity.

We now put the heuristic of the last section into a form which can be studied using the geometric structure of the graph. By the above, we want to restrict ourselves to using the expressions $\wr s \wr^2$ and $\wr sK \wr^2$. The task now is to express the polynomial decay of sK^n in terms of these two expressions. We first to give a heuristic study how one would do that, without trying to be mathematically rigorous (the formulas we get, of course, have been rigorously proven).

We will use the symbols $\phi = \wr s \wr^2$, and $\phi' = \wr Ks \wr^2 - \wr s \wr^2$. This way ϕ' denotes the change of ϕ under one step in our Markov chain. Our task now is similar to expressing the simple function $f(x) = cx^{-d}$ in terms of f and f', that is using a differential equation. Solving this problem one gets the expression $f^{d+1} = -c_2(f')^d$. Now, we want to apply this to ϕ, with the restriction that polynomial decay only occurs in the early phase. This means that our 'differential inequality' for ϕ should only imply polynomial decay for ϕ sufficiently large (or ϕ' sufficiently large compared to ϕ). This results in an inequality of the form:

$$\phi^{d+1} \leq c(-\phi' + \varepsilon\phi)^d.$$

which is almost exactly the Nash inequality (12). After this attempt at motivation, we quote the results of Diaconis and Saloff-Coste, translated to our setting:

Theorem 8.1 *Let K, π be a reversible Markov chain. Suppose that the Nash inequality*

$$\left[\wr s \wr^2\right]^{d+1} \leq c\left(\left[\wr s \wr^2 - \wr sK \wr^2\right] + N^{-1} \wr s \wr^2\right)^d \qquad (12)$$

holds for some constants $c, D > 0, N \geq 1$ and all distributions s. Then for any distribution s:

$$\wr sK^n \wr \leq \left(\frac{b}{n+1}\right)^{2d} \qquad \text{for } 0 \leq n \leq N$$

with $b = c(1 + (2N)^{-1})\lceil 1 + 4D \rceil$.

We omit the proof, which is fairly technical and can be found in [1]. This reference also contains a converse which shows that polynomial decay of the kernel K implies a Nash inequality. The combination of this decay bound and an eigenvalue bound yields bounds as in Theorem 14.1 below.

Chapter 3. Geometric techniques

This chapter will try to acquaint the reader with geometric concepts that are used to prove Nash inequalities. The two important concepts are those of moderate growth and of a local Poincaré inequality. For simplicity, this section only considers random walks on graphs with uniform stationary distribution. This means that the underlying graph is regular (see Section 5). Everything here has a version for the non-uniform case.

9 Bases of balls

Definition 9.1 *Consider a finite connected graph V. A* basis of balls *for V is a function $B : V \times \mathbb{Z}_{\geq 0} \to \mathcal{P}(V)$, which assigns to each vertex x ('center') and each nonnegative integer r ('radius') a subset of V ('ball'), with the following properties:*

1. *$B(x,0)=\{x\}$ for all x.*
2. *$B(x,r)$ is connected.*
3. *$r_1 < r_2 \Rightarrow B(x,r_1) \subset B(x,r_2)$.*
4. *There is a γ such that $B(x,\gamma) = V$ for all x. The minimal such γ is called the diameter of the graph.*

Note that any distance on a graph defines a basis of balls $B(x,r) := \{y \in V | d(x,y) \leq r\}$. In particular, the concept *basis of balls* is a generalization of this example with d being the shortest path distance. This generality is important for the purposes of this paper because the proof of the main theorem uses different (d_∞) balls. Diaconis and Saloff-Coste [1] provide arguments for traditional balls, but, as they note, their arguments work for general bases of balls.

10 Moderate growth

Moderate growth is a graph theoretic property which essentially measures the dimension of a given graph with respect to a collection of balls.

Definition 10.1 *A graph V has (A,d) moderate growth if*

$$|B(x,r)| \geq \frac{|V|}{A}\left(\frac{r+1}{\gamma}\right)^d \qquad \text{for all } x \in V \text{ and integers } r = 0, \ldots, \gamma. \tag{13}$$

that is, the volume of balls must be proportional to the d-th power of their radius. A thorough discussion of moderate growth can be found in [2].

11 Poincaré inequalities

The other geometric notion, Poincaré inequalities, measure the connectedness of a graph. There are two kinds of Poincaré inequalities: local and global. We start with the notion of global Poincaré inequalities. Using the inner product (4), for any reversible Markov chain K, π and any function $f : V \to \mathbb{R}$ we can write:

$$\begin{aligned}\mathcal{E}(f,f) &= < f, (I-K)f >= \sum_{x \in V} [f^2(x) - f(x)(Kf)(x)]\pi_x \\ &= \sum_{x,y \in V} [f^2(x) - K(x,y)f(y)f(x)]\pi_x = \frac{1}{2}\sum_{x,y \in V} [f(x) - f(y)]^2 K(x,y)\pi_x,\end{aligned}$$

where the last equality uses reversibility. Note that in this case the quantity $K(x,y)\pi_x$ is either 0 or a fixed constant, depending on whether there is an edge in the graph V between the vertices x, y.

The mean f_m of a function f with respect to the measure π is defined as:

$$f_m = \sum_{x \in V} f(x)\pi_x =< f_x, (1, 1, ..., 1) >$$

This is a constant, but we can also regard it as a constant function. Using this, we can write the variance of f as:

$$\|f - f_m\|^2 = \sum_{x \in V} [f(x) - f_m]^2 \pi_x = \sum_{x,y \in V} [f(x) - f(y)]^2 \pi_x \pi_y. \tag{14}$$

The use of this appears when one notes that the expression:

$$\lambda = \inf_f \frac{\mathcal{E}(f, f)}{\|f - f_m\|} \tag{15}$$

is the spectral gap for the matrix K (see, for example [4]), that is the difference of the two largest eigenvalues. In our case this is $1 - \beta_1$, where β_1 is the second-largest eigenvalue. Writing (15) in the inequality form we get a Poincaré inequality:

$$\|f - f_m\| \leq \frac{1}{\lambda} \mathcal{E}(f, f)$$

and λ is the largest constant for which this inequality holds for all functions f. This means that proving a Poincaré inequality for some constant gives an upper bound for the second largest eigenvalue. Up to constants, both the LHS and the RHS are sums of terms of the form $[f(x) - f(y)]^2$—the RHS is a sum over all x, y, the LHS is a sum over all x, y that are connected by an edge in the graph. Diaconis and Stroock [3] bound the difference $[f(x) - f(y)]^2$ for arbitrary x, y by differences along a path $\gamma_{x,y}$, and thus prove Poincaré inequalities and eigenvalue bounds in terms of the geometric properties of the graph (diameters and covering numbers).

12 The local Poincaré inequality

A local Poincaré inequality is a stronger version of a Poincaré inequality. First, we define 'smoothed' versions of a given function f, by replacing the value of f by the average over balls of radius r:

$$f_r := \frac{1}{|B(x, r)|} \sum_{y \in B(x,r)} f(y).$$

A local Poincaré inequality bounds the distance of f and f_r in terms of the Dirichlet form. A Markov chain satisfies a local Poincaré inequality with a parameter a if:

$$\|f, f_r\|^2 \leq a r^2 \mathcal{E}(f, f) \tag{16}$$

for all $r = 0, 1, ..., \gamma$, all functions f, and all vertices x. This is equivalent to saying that the random walks on the subgraphs $B(x, r)$, defined as in (3), satisfy a global Poincaré inequalities with parameter ar^2 for some fixed constant a. In the case $r = \gamma$, a local Poincaré inequality becomes a global one, and thus provides a bound on the second largest eigenvalue.

Local Poincaré inequalities are also proved by path arguments. The following result (in more generality) is due to Diaconis and Saloff-Coste [1]:

Theorem 12.1 *Suppose that we have a κ-regular graph, and collection of paths Γ, so that for each x, y, Γ contains a path connecting x with y. Then the Dirichlet form $\mathcal{E}(f,f)$ of the random walk on that graph satisfies the local Poincaré inequality*

$$\|f - f_r\|_2^2 \leq \eta(r)\mathcal{E}_Q(f,f)$$

for all functions f, with

$$\eta(r) = 2\kappa \max_{e \text{ edge}} \left\{ \sum_{\substack{\gamma_{x,y} \ni e \\ x,y \in B(*,r)}} \frac{|\gamma_{x,y}|}{|B(x,r)|} \right\} \tag{17}$$

where $x, y \in B(*,r)$ means that x, y are both contained in some basis ball of 'radius' r.

13 The least eigenvalue

The least eigenvalue β_{k-1} of a Markov chain describes the periodicity of the chain. A random walk on a bipartite graph will never get random, since the parity of n determines the position of the walk. In the spectrum of the kernel K this appears as an eigenvalue of -1. The natural graph on a set of lattice points without the self loops as in Figure 4 would be bipartite.

A graph is bipartite if and only if it has no cycles of odd length. Diaconis and Stroock [3] use cycles of odd length to prove a bound on the least eigenvalue. The proof is similar to the way they prove a Poincaré inequality. We quote the uniform version of their theorem:

Theorem 13.1 *Let K be a Markov kernel of a random walk on a κ-regular graph. For each vertex x let $\gamma_{x,x}$ be a cycle of odd length containing x. Then the least eigenvalue β_{k-1} satisfies:*

$$\frac{1}{1+\beta_{k-1}} \leq 2\kappa \max_{e \text{ edge}} \left\{ \sum_{\gamma_{x,x} \ni e} |\gamma_{x,x}| \right\}.$$

The least eigenvalue does not play a very important role in most cases, because it is usually much smaller in absolute value than β_1. This is true in our case as well—we can get a bound on β_{k-1} without much difficulty.

14 Convergence bounds

Diaconis and Saloff-Coste [1] prove that local Poincaré inequality, moderate growth, and a bound on the least eigenvalue together imply a Nash inequality and a convergence bound:

Theorem 14.1 *Let K, π be a random walk on a regular graph V. Assume that (K, π) has moderate growth (13) and satisfies a local Poincaré inequality (16). Assume further that the least eigenvalue β_{k-1} satisfies $\beta_{k-1} \geq -1 + \frac{1}{a\gamma^2}$. Then*

$$\|(K_x^n/\pi) - 1\|_2 \leq a_1 e^{-m/(a\gamma^2)} \quad \textit{for} \quad n = 2a\gamma^2 + m + 1$$

with $m \geq 0$ and $a_1 = (2e(1+d)A)^{\frac{1}{2}}(2+d)^{d/4}$.

We omit the proof of this theorem, which can be found in [1] It combines the polynomial bound implied by Nash inequalities for the early phase of the walk, and the exponential eigenvalue bound for the later phase. It is the most important result needed for proving the main theorem of this paper.

Chapter 4. The main theorem

In this chapter we will prove that the random walk defined in (3) has moderate growth and satisfies a local Poincaré inequality. This will imply the desired convergence.

Convexity is heavily used in the following sections. Recall that if A and B are convex subsets of $\mathbb{R}^d$, then $A+B$ is convex. If A contains the origin, and $0 \le \lambda \le 1$, then $\lambda A \subset A$. The following identity will prove to be a useful tool. For any finite set of lattice points V and the hypercube $\mathcal{D}$ defined in Section 1:

$$\mathrm{vol}(V+\mathcal{D}) = |V|. \tag{18}$$

Let $B(x,r)$ be the set of points in V of d_∞-distance at most r from x. Then $B(x,r)$ forms a basis of balls as in Definition (9.1). We will use this collection of balls throughout this chapter. In the next section we will use the volume expression 18 and set inclusion arguments in $\mathbb{R}^d$ to estimate the volume growth of our graph, that is $|B(x,r)|$ as a function of r.

15 Moderate growth

The following lemma shows that V satisfies a moderate growth criterion: the number of points of V inside a d_∞ ball grows rapidly as r increases. We will use a contraction in $\mathbb{R}^d$ and the fact that the number of lattice points in a region is bounded by the volume of a slightly larger region (see (18)). We use ε-arguments to get better constants (eg. compare the number of lattice points inside $D(1)$ and $D(1-\varepsilon)$).

Lemma 15.1 *Any convex subset V of the lattice $\mathbb{Z}^d$ has moderate growth*

$$|B(x,r)| \ge \frac{|V|}{A}\left(\frac{r+1}{\gamma_\infty}\right)^d \quad \textit{for all } x \in V \textit{ and integers } r = 0,1,\ldots,\gamma_\infty, \tag{19}$$

where γ_∞ is the d_∞ diameter of the set V, and $A = 5^d$.

Proof. For convenient description of contractions we will assume that x is the origin. By Lemma 2.2 there is a set C and $\varepsilon > 0$ such that V is the set of lattice points in the $\frac{1}{2}-\varepsilon$ neighborhood of C, and there are no additional lattice points in the $\frac{1}{2}$-negihborhood. We can write this as $V = \mathbb{Z}^d \cap (C + D(\frac{1}{2}-\varepsilon)) = \mathbb{Z}^d \cap (C+\mathcal{D})$.

Let $C' := C + D(\frac{1}{2}-\varepsilon)$, and note that C' is also convex. By the above know that $V \subset C'$. Also, $V \subset D(\gamma_\infty)$, where $D(\gamma_\infty)$ is the closed d_∞-ball of radius γ_∞. Let $r+1 \le \gamma_\infty$, then the above two relations imply

$$\frac{r+1}{\gamma_\infty} V \subset \frac{r+1}{\gamma_\infty}\left(C' \cap D(\gamma_\infty)\right) \subset C' \cap D(r+1).$$

The second inclusion follows from the convexity of C'. Now let y be in the RHS, then there is a $y' \in C$ with $d_\infty(y',y) \le \frac{1}{2}-\varepsilon$. Then by the definition of convexity there is a $y'' \in V$ such that $d_\infty(y'',y') \le \frac{1}{2}$. This means that

$$d_\infty(y'', x) \le d_\infty(y'', y') + d_\infty(y', y) + d_\infty(y, x) \le r + 2 - \varepsilon.$$

Since x and y'' have integer distance, this implies $d(x, y'') \le r + 1$, and so $y'' \in B(x, r + 1)$. But $d(y, y'') \le 1 - \varepsilon$ implies $y \in B(x, r + 1) + D(1) \subset B(x, r) + 4\mathcal{D}$. So we have proved

$$\frac{r+1}{\gamma_\infty} V \subset B(x, r) + 4\mathcal{D}.$$

Using this, we can write

$$\frac{r+1}{\gamma_\infty}(V + \mathcal{D}) \subset \frac{r+1}{\gamma_\infty} V + \mathcal{D} \subset B(x, r) + 5\mathcal{D} = \bigcup_{y \in B(x,r)} (y + 5\mathcal{D}).$$

The volumes of these sets are related to the numbers of lattice points they contain. The LHS has volume exactly $(\frac{r+1}{\gamma_\infty})^d |V|$, and where the volume of the RHS is bounded above by $5^d |B(x, r)|$. The inequality between these volumes implies the stated moderate growth criterion. This proof works for $r \le \gamma_\infty - 1$, and the $r = \gamma_\infty$ case is straightforward. □

The following lemma bounds the number of points in a ball of radius $2r$ in terms of the number of points in a ball of radius r in V. It will be needed in a later section. The proof goes along the same lines as the last one.

Lemma 15.2 $|B(x, 2r)| \le 6^d |B(x, r)|$

Proof. Using the notation of the last lemma, we have $B(x, 2r) \subset C' \cap D(x, 2r)$. Therefore $\frac{1}{2} B(x, 2r) \subset \frac{1}{2}(C' \cap D(x, 2r)) \subset C' \cap D(x, 2r)$ by convexity. Now if y is in RHS, then there is a point $y' \in V$ such that $d(y, y') < 1 - \varepsilon$. This implies $y' \in B(x, r)$, and so $y \in B(x, r) + 2\mathcal{D}$. Therefore:

$$\frac{1}{2} B(x, 2r) + \mathcal{D} \subset B(x, r) + 3\mathcal{D}.$$

The desired inequality follows from comparing the volumes. □

16 Local Poincaré inequality

Poincaré and local Poincaré inequalities are traditionally proved by path arguments. The main part of the following proof consists of a fairly technical argument estimating the number of almost straight geodesic paths $\gamma_{x,y}$ of bounded length containing a given edge e. For fixed x, the fact that paths are almost straight forces the number of possible y-s to be in a d-dimensional cone, which we will estimate by a prism.

Lemma 16.1 *The random walk* (3) *on the convex set of lattice points V satisfies a local Poincaré inequality*

$$\|f - f_r\|^2 \le a r^2 \mathcal{E}(f, f) \tag{20}$$

with $a = \frac{3}{14} d^3 42^d$

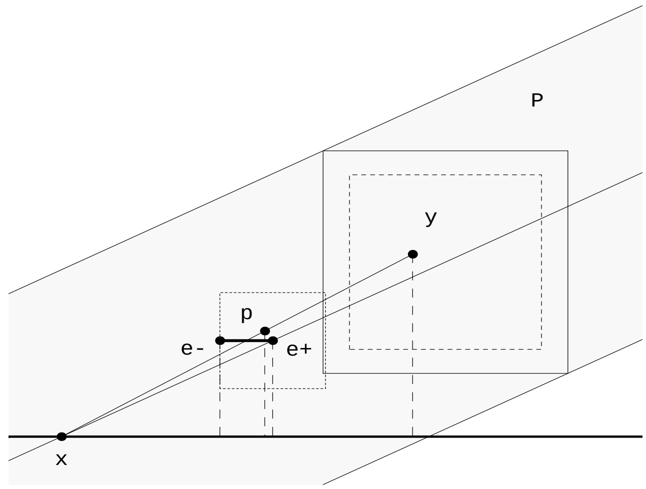

Figure 7: All possible points y are contained in a *d-dimensional prism*

Proof. We will use the paths constructed in Section 4, and Theorem 12.1. Using that we have to bound

$$\eta(r) = 4d \max_{e \text{ edge}} \left\{ \sum_{\substack{\gamma_{x,y} \ni e \\ x,y \in B(*,r)}} \frac{|\gamma_{x,y}|}{|B(x,r)|} \right\}$$

by an expression of the form ar^2. Here $x, y \in B(*, r)$ means that x, y are both contained in some basis ball of 'radius' r.

First we need to count the number of paths containing a given edge e. Let $\gamma_{x,y}$ be such a path with y being closer to e, that is $d_\infty(x, e_-) \geq d_\infty(y, e+)$. We fix x, and for the sake of simple notation assume that it is in the origin. Also, WLOG we can assume that the coordinates $(e_1, ..., e_d)$ of e_+ are nonnegative, and e_1 is maximal. Then y has coordinates so that $|y_i| \leq 2e_1$, and since our paths are geodesic, we have $x_1 \leq e_1 \leq y_1 \leq r$. These imply that the first coordinate of y satisfies

$$e_1 \leq y_1 \leq e_1 + r/2. \tag{21}$$

(step 1) We constructed our paths so that the line $\overline{xy}$ must intersect the unit d_∞ ball about e_+, say in a point p, where $e_1 - 1 \leq p_1$. Then, y is contained in a dilation of this ball from the origin by a factor y_1/p_1 (see Figure 7).

This means that the d_∞-distance of y from the line $\overline{xe_+}$ is at most $y_1/p_1 \leq 2e_1/(e_1 - 1)$. Thus the distance of the points in the cubicle $y + \mathcal{D}$ from $\overline{xe_+}$ is at most $2e_1/(e_1 - 1) + \frac{1}{2}$. For $e_1 \geq 3$, this is not more than 3.5. Therefore $y + \mathcal{D} \subset P$, where $P = \overline{xe_+} + D(3.5)$, a d-dimensional infinite prism, geometrically the trace of translating the hypercube $D(3.5)$ keeping its center on the line $\overline{xe_+}$.

(step 2) For the next part of the proof we will need the following two facts from geometry (the proofs are given later):

(22) The projection of a d-dimensional unit hypercube to a *(d-1)*-dimensional hyperplane has volume ("area") at most $\sqrt{d}$.

(23) Let P be a d-dimensional prism with axis parallel to the unit vector $\vec{a}$. For any vector $\vec{v}$, let $P(\vec{v})$ be the volume of the cross section of P by a *(d-1)*-dimensional hyperplane with normal vector $\vec{v}$. If $\vec{v}$ is a unit vector, then $P(\vec{a})/P(\vec{v}) = \vec{a} \cdot \vec{v}$ (a dot product).

Consider the prism $P = \overline{xe_+} + D(3.5)$. Its cross-section by a hyperplane with normal vector $x\vec{e}_+$ is a projection of the hypercube $D(3.5)$. Thus, by (22), the volume of the cross-section is at most $\sqrt{d}7^{d-1}$.

Let $\vec{a}$ be the unit vector parallel to $x\vec{e}_+$, and $\vec{\vartheta}_1$ be the first unit coordinate vector. Note that we assumed that e_1 is the largest coordinate of $x\vec{e}_+$. This means that the dot product $\vec{\vartheta}_1 \cdot \vec{a}$ is bounded below by $\sqrt{1/d}$. Thus by (23) the volume $P(\vec{\vartheta}_1) \leq \sqrt{d}P(\vec{a}) \leq d7^{d-1}$.

(step 3)

Using the result of step 1 and (21), we know that any cubicle $y + \mathcal{D}$ is contained in the intersection of the prism P and the region $\{r \in \mathbb{R}^d | e_1 - \frac{1}{2} \leq z_1 \leq e_1 + \frac{r}{2} + \frac{1}{2}\}$. Any eligible candidate for y occupies a cubicle of unit volume in this region. The volume of this region can be computed as height×base, that is $(\frac{r}{2} + 1) \times d7^{d-1}$. Therefore the number of possible points y is at most $(\frac{r}{2} + 1)d7^{d-1}$ for $e_1 \geq 3$. A simple argument shows that this bound also works for $e_1 \leq 2$. This implies that the number of paths containing e is bounded above by

$$d7^{d-1}(r/2 + 1)\max(|B(e_-, r)|, |B(e_+, r)|) \leq d7^{d-1}(r/2 + 1)|B(x, 2r)|$$

From Lemma 15.2 we know that $|B(x, 2r)| \leq 6^d|B(x, r)|$. Since $\gamma_{x,y}$ is a geodesic path inside a ball $B(*, r)$, the length $|\gamma_{x,y}|$ is bounded by dr. Using this gives the bound $\eta(r) \leq \frac{1}{14}d^3 42^d r(r + 2)$, which is at most $\frac{3}{14}d^3 42^d r^2$ for $r \geq 1$. □

Proof of the facts (22) and (23). First note the following intuitive fact (the proof can be found in [5]). Let $H(\vec{u})$ and $H(\vec{v})$ be hyperplanes with unit normal vectors $\vec{u}, \vec{v}$. Let T be a polytope in $H(\vec{v})$, and let $\pi(T)$ be the orthogonal projection of T to $H(\vec{u})$. Then the *(d-1)*-dimensional volumes satisfy

$$\mathrm{vol}(\pi(T))/\mathrm{vol}(T) = \vec{u} \cdot \vec{v}. \tag{24}$$

The fact (23) immediately follows from this expression when one notices that the perpendicular cross-section of an infinite prism can be thought of as an orthogonal projection of any other cross-section.

For (22), let $\mathcal{D}$ be a d-dimensional unit hypercube as defined in (1). Let p be a plane with normal vector $\vec{v}$. WLOG we can assume that all coordinates of $\vec{v}$ are nonnegative. Now each line parallel to $\vec{v}$ that intersects the hypercube $\mathcal{D}$ must intersect one of the d faces (that is, *(d-1)*-dimensional components) that are in the hyperplanes $\xi_i = \frac{1}{2}$. Therefore the projection of the hypercube is the same as the projection of the union of these faces. Each face is a *(d-1)*-dimensional hypercube of area 1. Thus by (24), the area of the projection of such a face is given by $\vec{\vartheta}_i \cdot a$, where $\vec{\vartheta}_i$ is the i-th coordinate vector. Thus the maximal area of the projection is given by:

$$\sup_{\vec{v}} \left\{ \sum_{i=1}^{d} = \vec{\vartheta}_i \cdot \vec{v} \right\}$$

The supremum is taken over all unit vectors $\vec{v}$. Note that the sum on the RHS is just the sum of the coordinates of the unit vector $\vec{v}$, which is bounded by $\sqrt{d}$ by the inequality between the arithmetic mean and the root-mean-square. □

17 A bound on the least eigenvalue

The main theorem requires a bound on the least eigenvalue. For each x, let $\gamma_{x,x}$ be a cycle of vertices whose first coordinates are $(x_1, x_1+1, x_1+2, ..., x_1+i, x_1+i, x_1+i-1, ..., x_1)$, and the rest of the coordinates are the same as those of x. Here the lattice point $(x_1+i, x_2, ..., x_d)$ is on the boundary of V, that is it has a self-loop attached to it. In other words, the loop $\gamma_{x,x}$ goes straight to the boundary, uses the loop there and then returns to x. This way each edge is used in at most γ_∞ times by the $\gamma_{x,x}$-s, since there are at most γ_∞ vertices with the same first coordinate. Also, each cycle has length at most $2\gamma_\infty + 1 \leq 3\gamma_\infty$. Therefore by Theorem 13 we get the bound

$$\beta_{k-1} \geq -1 + \frac{1}{4d \times 3\gamma_\infty^2} \geq -1 + \frac{1}{a\gamma_\infty^2} \tag{25}$$

where a is the constant of the local Poincaré inequality (20).

18 Bound on the convergence rate

The moderate growth criterion (19), the local Poincaré inequality (20), and the least eigenvalue bound (25) together imply a bound on the convergence rate by Theorem 14.1. Thus we have proven the theorem:

Theorem 18.1 *Let V be a finite convex subset of the lattice $\mathbb{Z}^d$. Then the random walk (3) on V satisfies the convergence bound:*

$$||K_x^n/\pi - 1||_2 \leq a_1 e^{-m/(a\gamma_\infty^2)} \quad \text{for} \quad n = 2a\gamma_\infty^2 + m + 1 \tag{26}$$

with $m \geq 0$, $a_1 = (2e(1+d)5^d)^{\frac{1}{2}}(2+d)^{d/4}$, and $a = \frac{3}{14}d^3 42^d$.

In particular, this means that for fixed $\varepsilon > 0$, it takes at most $c\gamma^2$ steps for the distance (26) to get less then ε, where γ is the graph-theoretic diameter ($\gamma \geq \gamma_\infty$). Here the constant c depends only on ε and the dimension d. The time it takes for the child to be in any exhibit hall with about equal probability is proportional to the diameter-square of the museum. By now the parents have seen enough of the exhibit 'Cubism and Mathematics'—it is time to go.

References

[1] Diaconis P. and Saloff-Coste L. (1995) *Nash inequalities for finite Markov chains.* To Appear.

[2] Diaconis P. and Saloff-Coste L. (1994) *Moderate growth and random walk on finite groups.* Geom. and Funct. Anal., 4, 1-34.

[3] Diaconis P. and Stroock D. (1991) *Geometric bounds for eigenvalues of Markov chains.* Ann. Appl. Prob. 1, 36-61.

[4] Horn R. and Johnson C. (1985) *Matrix analysis.* Cambridge University Press, New York.

[5] Sommerville, D. M. Y. (1958) *An introduction to the geometry of n dimensions.* Dover Publications, New York.