Lectures on finite Markov chains

Laurent Saloff-Coste
CNRS & Université Paul Sabatier, UMR 55830

École d'été de probabilités de St Flour 1996

Contents

Chapter 1

Introduction and background material

1.1 Introduction

I would probably never have worked on finite Markov chains if I had not met Persi Diaconis. These notes are based on our joint work and owe a lot to his broad knowledge of the subject although the presentation of the material would have been quite different if he had given these lectures.

The aim of these notes is to show how functional analysis techniques and geometric ideas can be helpful in studying finite Markov chains from a quantitative point of view.

A Markov chain will be viewed as a Markov operator K acting on functions defined on the state space. The action of K on the spaces $\ell^p(\pi)$ where π is the stationary measure of K will be used as an important tool. In particular, the Hilbert space $\ell^2(\pi)$ and the Dirichlet form

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y} |f(x)-f(y)|^2 K(x,y)\pi(x)$$

associated to K will play crucial roles. Functional inequalities such as Poincaré inequalities, Sobolev and Nash inequalities, or Logarithmic Sobolev inequalities will be used to study the behavior of the chain.

There is a natural graph structure associated to any finite Markov chain K. The geometry of this graph and the combinatorics of paths enter the game as tools to prove functional inequalities such as Poincaré or Nash inequalities and also to study the behavior of different chains through comparison of their Dirichlet forms.

The potential reader should be aware that these notes contain no probabilistic argument. Coupling and strong stationary times are two powerful techniques that have also been used to study Markov chains. They form a set of techniques that are very different in spirit from the one presented here. See, e.g., [1, 19]. Diaconis' book [17] contains a chapter on these techniques. David Aldous and Jim Fill are writing a book on finite Markov chains [3] that contains many wonderful things.

The tools and ideas presented in these notes have emerged recently as useful techniques to obtain quantitative convergence results for complex finite Markov chains. I have tried to illustrate these techniques by natural, simple but non trivial examples. More complex (and more interesting) examples require too much additional specific material to be treated in these notes. Here are a few references containing compelling examples:

– For eigenvalue estimates using path techniques, see [35, 41, 53, 72].
– For comparison techniques, see [23, 24, 30]
– For other geometric techniques, see [21, 38, 39, 43, 60].

Acknowledgements: Many thanks to Michel Benaim, Sergei Bobkov, Persi Diaconis, Susan Holmes, Michel Ledoux, Pascal Lezaud and Laurent Miclo for their help. Thanks also to David Aldous, Jim Fill, Mark Jerrum, Alistair Sinclair for useful discussions and comments over the years.

1.1.1 My own introduction to finite Markov chains

Finite Markov chains provide nice exercises in linear algebra and elementary probability theory. For instance, they can serve to illustrate diagonalization or triangularization in linear algebra and the notion of conditional probability or stopping times in probability. That is often how the subject is known to professional mathematicians.

The ultimate results then appear to be the classification of the states and, in the ergodic case, the existence of an invariant measure and the convergence of the chain towards its invariant measure at an exponentiel rate (the Perron-Frobenius theorem). Indeed, this set of results describes well the asymptotic behavior of the chain.

I used to think that way, until I heard Persi Diaconis give a couple of talks on card shuffling and other examples.

How many times do you have to shuffle a deck of cards so that the deck is well mixed?

The fact that shuffling many, many times does mix (the Perron-Frobenius Theorem) is reassuring but does not at all answer the question above.

Around the same time I started to read a paper by David Aldous [1] on the subject because a friend of mine, a student at MIT, was asking me questions about it. I was working on analysis on Lie groups and random walk on finitely generated, infinite group under the guidance of Nicolas Varopoulos. I had the vague feeling that the techniques that Varopolous had taught me could also be applied to random walks on finite groups. Of course, I had trouble deciding whether this feeling was correct or not because, on a finite set, everything is always true, any functional inequality is satisfied with appropriate constants.

Consider an infinite group G, generated by a finite symmetric set S. The associated random walk proceeds by picking an element s in S at random and move from the current state x to xs. An important nontrivial result in random walk theory is that the transient/recurrent behavior of these walks depends only on G and not on the choosen generating set S. The proof proceeds by comparison of Dirichlet forms. The Dirichlet form associated to S is

$$\mathcal{E}_S(f,f) = \frac{1}{2|S|} \sum_{g\in G, h\in S} |f(g) - f(gh)|^2.$$

If S and T are two generating sets, one easily shows that there are constants $a, A > 0$ such that

$$a\mathcal{E}_S \leq \mathcal{E}_T \leq A\mathcal{E}_S.$$

To prove these inequalities one writes the elements of S as finite products of elements of T and vice versa. They can be used to show that the behavior of finitely generated symmetric random walks on G, in many respects, depends only on G, not on the generating set.

I felt that this should have a meaning on finite groups too although clearly, on a finite group, different generating finite sets may produce different behaviors.

I went to see Persi Diaconis and we had the following conversation:

L: Do you have an example of finite group on which there are many different walks of interest?
P: Yes, the symmetric group S_n!
L: Is there a walk that you really know well?
P: Yes there is. I know a lot about random transpositions.
L: Now, we need another walk that you do not know as well as you wish.
P: Take the generators $\tau = (1,2)$ and $c^{\pm 1} = (1,\ldots,n)^{\pm 1}$.
L& P: Lets try it. Any transposition can be written as a product of τ and $c^{\pm 1}$ of length at most $10n$. Each of τ, c, c^{-1} is used at most $10n$ times to write a given transposition. Hence, (after some computations) we get

$$\mathcal{E}_T \leq 100\, n^2\, \mathcal{E}_S$$

where $\mathcal{E}_T$ is the Dirichlet form for random transpositions and $S = \{\tau, c, c^{-1}\}$. What can we do with this? Well, the first nontrivial eigenvalue of random transpositions is $1 - 2/n$ by Fourier analysis. This yields a bound of order $1 - 50/n^3$ for the walk based on the generating set S.
L: I have no idea whether this is good or not.
P: Well, I do not know how to get this result any other way (as we later realized $1 - c/n^3$ is the right order of magnitude for the first nontrivial eigenvalue of the walk based on S).
L: Do you have any other example? ....

This took place during the spring of 1991. The conversation is still going on and these notes are based on it.

1.1.2 Who cares?

There are many ways in which finite Markov chains appear as interesting or useful objects. This section presents briefly some of the aspects that I find most compelling.

Random walks on finite groups. I started working on finite Markov chains by looking at random walks on finite groups. This is still one of my favorite aspects of the subject. Given a finite group G and a generating set $S \subset G$, define a Markov chain as follows. If the current state is g, pick s in S uniformly at random and move to gs. For instance, take $G = S_n$ and $S = \{\mathrm{id}\} \cup \{(i,j) : 1 \leq i < j \leq n\}$. This yields the "random transpositions" walk. Which generating sets of S_n are most efficient? Which sets yield random walks that are slow to converge? How slow can it be? More generally, which groups carry fast generating sets of small cardinality? How does the behavior of random walks relate to the algebraic structure of the group? These are some of the questions that one can ask in this context. These notes do not study finite random walks on groups in detail except for a few examples. The book [17] gives an introduction and develops tools from Fourier analysis and probability theory. See also [42]. The survey paper [27] is devoted to random walks on finite groups. It contains pointers to the literature and some open questions. Many examples of walks on the symmetric group are treated by comparison with random transpositions in [24]. M. Hildebrand [49] studies random transvections in finite linear groups by Fourier analysis. The recent paper of D. Gluck [45] contains results for some classical finite groups that are based on the classification of simple finite groups. Walks on finite nilpotent groups are studied in [25, 26] and in [74, 75, 76].

Markov Chain Monte Carlo. Markov chain Monte Carlo algorithms use a Markov chain to draw from a given distribution π on a state space $\mathcal{X}$ or to approximate π and compute quantities such as $\pi(f)$ for certain functions f. The **Metropolis** algorithm and its variants provide ways of constructing Markov chains which have the desired distribution π as stationary measure. For instance let Λ be a 100 by 100 square grid, $\mathcal{X} = \{x : \Lambda \to \{\pm 1\}\}$ and

$$\pi(x) = z(c)^{-1} \exp\left\{ c \left(\sum_{i,j: i \sim j} x_i x_j + h \sum_i x_i \right) \right\}$$

where $z(c)$ is the unknown normalizing constant. This is the **Gibbs** measure of a finite two-dimentional Ising model with inverse temperature $c > 0$ and external field strength h. In this case the Metropolis chain proceed as follows. Pick a site $i \in \Lambda$ at random and propose the move $x \to x^i$ where x^i is obtained from x by changing $x(i)$ to $-x(i)$. If $\pi(x^i)/\pi(x) \geq 1$ accept this move. If not, flip a coin with probability of heads $\pi(x^i)/\pi(x)$. If the coin comes up heads, move to x^i. If the coins comes up tails, stay at x. It is not difficult to show that this chain has stationary measure π as desired. It can then be used (in principle) to draw from π (i.e., to produce typical configurations), or to estimate the normalizing constant $z(c)$. Observe that running this chain implies computing $\pi(x^i)/\pi(x)$. This is reasonable because the unknown normalizing constant disappears in this ratio and the computation only involves looking at neighbors of the site i.

Application of the Metropolis algorithm are widespread. Diaconis recommends looking at papers in the Journal of the Royal Statistical Society, Series B, 55(3), (1993) for examples and pointers to the literature. Clearly, to validate (from a theoretical point of view) the use of this type of algorithm one needs to be able to answer the question: how many steps are sufficient (necessary) for the chain to yield a good approximation of π? These chains and algorithms are often used without any theoretical knowledge of how long they should be run. Instead, the user most often relies on experimental knowledge, hoping for the best.

Let us emphasize here the difficulties that one encounters in trying to produce theoretical results that bear on applications. In order to be directly relevant to applied work, theoretical results concerning finite Markov chains must not only be quantitative but they must yield bounds that are close to be sharp. If the bounds are not sharp enough, the potential user is likely to disregard them as unreasonably conservative (and too expensive in running time). It turns out that many finite Markov chains are very effective (i.e., are fast to reach stationarity) for reasons that seem to defy naive analysis. A good example is given by the Swendsen-Wang algorithm which is a popular sampling procedure for Ising configuration according to the Gibbs distribution [77]. This algorithm appears to work extremely well but there are no quantitative theoretical results to support this experimental finding. A better understood example of this phenomenon is given by random transpositions (and other walks) on the symmetric group. In this case, a precise analysis can be obtained through the well developed representation theory of the symmetric group. See [17].

Theoretical Computer Science. Much recent progress in quantitative finite Markov chain theory is due to the Computer Science community. I refer the reader to [54, 56, 71, 72] and also [31] for pointers to this literature. Computer scientists are interested in classifying various combinatorial tasks according to their complexity. For instance, given a bipartite connected graph on $2n$ vertices with vertex set $O \cup I$, $\#O = \#I = n$, and edges going from I to O, they ask whether or not there exists a deterministic algorithm in polynomial time in n for the following tasks:

(1) decide whether there exists a perfect matching in this graph

(2) count how many perfect matchings there are.

A perfect matching is a set of n edges such that each vertex appears once. It turn out that the answer is yes for (1) and most probably no for (2) in a precise sense, that is, (2) is an example of a # P-complete problem. See e.g., [72].

Using previous work of Broder, Mark Jerrum and Alistair Sinclair were able to produce a stochastic algorithm which approximate the number of matchings in polynomial time (for a large class of graphs). The main step of their proof consists in studying a finite Markov chain on perfect and near perfect matchings. They need to show that this chain converges to stationarity in polynomial time. They introduce paths and their combinatorics as a tool to solve this problem. See [54, 72]. This technique will be discussed in detail in these notes.

Computer scientists have a host of problems of this type, including the celebrated problem of approximating the volume of a convex set in high dimension. See [38, 39, 56, 60].

To conclude this section I would like to emphasize that although the present notes only contain theoretical results these results are motivated by the question obviously relevant to applied works:

How many steps are needed for a given finite Markov chain to be close to equilibrium?

1.1.3 A simple open problem

I would like to finish this introduction with a simple example of a family of Markov chains for which the asymptotic theory is trivial but satisfactory quantitative results are still lacking. This example was pointed out to me by M. Jerrum.

Start with the hypercube $\mathcal{X} = \{0,1\}^n$ endowed with its natural graph structure where x and y are neighbors if and only if they differ at exactly one coordinate, that is, $|x - y| = \sum |x_i - y_i| = 1$. The simple random walk on this graph can be analysed by commutative Fourier analysis on the group $\{0,1\}^n$ (or otherwise). The corresponding Markov operator has eigenvalues $1 - 2j/n$, $j = 0, 1, \ldots, n$, each with multiplicity $\binom{n}{j}$. It can be shown that this walk reaches approximate equilibrium after $\frac{1}{4} n \log n$ many steps in a precise sense.

Now, fix a sequence $\mathbf{a} = (a_i)_1^n$ of non-negative numbers and $b > 0$. Consider

$$\mathcal{X}(\mathbf{a}, b) = \left\{ x \in \{0,1\}^n : \sum a_i x_i \leq b \right\}.$$

This is the hypercube chopped by a hyperplane. Consider the chain $K = K_{\mathbf{a},b}$ on this set defined by $K(x,y) = 1/n$ if $|x - y| = 1$, $K(x,y) = 0$ if $|x - y| > 1$ and $K(x,x) = 1 - n(x)/n$ where $n(x) = n_{\mathbf{a},b}(x)$ is the number of y in $\mathcal{X}(\mathbf{a}, b)$ such that $|x - y| = 1$. This chain has the uniform distribution on $\mathcal{X}(\mathbf{a}, b)$ as stationary measure.

At this writing it is an open problem to prove that this chain is close to stationarity after $n^{O(1)}$ many steps, uniformly over all choices of $\mathbf{a}, b$. A partial result when the set $\mathcal{X}(\mathbf{a}, b)$ is large enough will be described in these notes. See also [38].

1.2 The Perron-Frobenius Theorem

One possible approach for studying finite Markov chains is to reduce everything to manipulations of finite-dimensional matrices. Kemeny and Snell [57] is a useful reference written in this spirit. From this point of view, the most basic result concerning the asymptotic behavior of finite Markov chains is a theorem in linear algebra, namely the celebrated Perron-Frobenius theorem.

1.2.1 Two proofs of the Perron-Frobenius theorem

A *stochastic matrix* is a square matrix with nonnegative entries whose rows all sum to 1.

Theorem 1.2.1 *Let M be an n-dimensional stochastic matrix. Assume that there exists k such that M^k has all its entries positive. Then there exists a row vector $m = (m_j)_1^n$ with positive entries summing to 1 such that for each $1 \le i \le n$,*

$$\lim_{\ell \to \infty} M^{\ell}_{i,j} = m_j. \tag{1.2.1}$$

Furthermore, $m = (m_i)_1^n$ is the unique row vector such that $\sum_1^n m_i = 1$ and $mM = m$.

We start with the following Lemma.

Lemma 1.2.2 *Let M be an n-dimensional stochastic matrix. Assume that for each pair $(i,j), 1 \le i,j \le n$ there exists $k = k(i,j)$ such that $M^k_{i,j} > 0$. Then there exists a unique row vector $m = (m_j)_1^n$ with positive entries summing to 1 such that $mM = m$. Furthermore, 1 is a simple root of the characteristic polynomial of M.*

Proof: By hypothesis, the column vector $\mathbf{1}$ with all entries equal to 1 satisfies $M\mathbf{1} = \mathbf{1}$. By linear algebra, the transpose M^t of M also has 1 as an eigenvalue, i.e., there exists a row vector v such that $vM = v$. We claim that $|v|$ also satisfies $|v|M = |v|$. Indeed, we have $\sum_i |v_i| M_{i,j} \ge |v_j|$. If $|v|M \ne |v|$, there exists j_0 such that $\sum_i |v_i| M_{i,j_0} > |v_{j_0}|$. Hence, $\sum_i |v_i| = \sum_j \sum_i |v_i| M_{i,j} > \sum_j |v_j|$, a contradiction. Set $m_j = v_j / (\sum_i |v_i|)$. The weak irreducibility hypothesis in the lemma suffices to insure that there exists ℓ such that $A = (I + M)^{\ell}$ has all its entries positive. Now, $mA = 2^{\ell} m$ implies that m has positive entries.

Let u be such that $uM = u$. Since $|u|$ is also an eigenvector its follows that the vector u^+ with entries $u_i^+ = \max\{u_i, 0\}$ is either trivial or an eigenvector. Hence, u^+ is either trivial or equal to u (because it must have positive entries). We thus obtain that each vector $u \ne 0$ satisfying $uM = u$ has entries that are either all positive or all negative. Now, if m, m' are two normalized eigenvectors with positive entries then $m - m'$ is either trivial or an eigenvector. If $m - m'$ is not trivial its entries must change sign, a contradiction. So, in fact, $m = m'$.

To see that 1 has geometric multiplicity one, let V be the space of column vectors. The subspace $V_0 = \{v : \sum_i v_i = 0\}$ is stable under M: $MV_0 \subset V_0$ and $V = \mathbb{R}\mathbf{1} \oplus V_0$. So either $M - I$ is invertible on V_0 or there is a $0 \ne v \in V_0$ such that $Mv = v$. The second possibility must be ruled out because we have shown that the entries of such a v would have constant sign. This ends the proof of Lemma 1.2.2. We now complete the proof of Theorem 1.2.1 in two different ways.

PROOF (1) OF THEOREM 1.2.1: Using the strong irreducibility hypothesis of the theorem, let k be such that $\forall\ i,j \quad M^k_{i,j} > 0$. Let $m = (m_i)_1^n$ be the row vector constructed above and set $M^\infty_{i,j} = m_j$ so that M^∞ is the matrix with all rows equal to m. Observe that

$$MM^\infty = M^\infty M = M^\infty \tag{1.2.2}$$

and that $M^k_{i,j} \geq cM^\infty_{i,j}$ with $c = \min_{i,j}\{M^k_{i,j}/M^\infty_{i,j}\} > 0$. Consider the matrix

$$N = \frac{1}{1-c}\left(M^k - cM^\infty\right)$$

with the convention that $N = 0$ if $c = 1$ (in which case we must indeed have $M^k = M^\infty$). If $0 < c < 1$, N is a stochastic matrix and $NM^\infty = M^\infty N = M^\infty$. In all cases, the entries of $(N - M^\infty)^\ell = N^\ell - M^\infty$ are bounded by 1, in absolute value, for all $\ell = 1, 2, \ldots$. Furthermore

$$\begin{aligned} M^k - M^\infty &= (1-c)(N - M^\infty) \\ M^{k\ell} - M^\infty &= (M^k - M^\infty)^\ell = (1-c)^\ell (N - M^\infty)^\ell. \end{aligned}$$

Thus

$$|M^{k\ell}_{i,j} - M^\infty_{i,j}| \leq (1-c)^\ell.$$

Consider the norm $\|A\|_\infty = \max_{i,j} |A_{i,j}|$ on matrices. The function

$$\ell \to \|M^\ell - M^\infty\|_\infty$$

is nonincreasing because $M^{\ell+1} - M^\infty = M(M^\ell - M^\infty)$ implies

$$\begin{aligned} (M^{\ell+1} - M^\infty)_{i,j} &= \sum_s M_{i,s}(M^\ell - M^\infty)_{s,j} \\ &\leq \left(\sum_s M_{i,s}\right) \|M^\ell - M^\infty\|_\infty = \|M^\ell - M^\infty\|_\infty. \end{aligned}$$

Hence,

$$\max_{i,j}\left\{|M^\ell_{i,j} - m_j|\right\} \leq (1-c)^{\lfloor \ell/k \rfloor}.$$

In particular $\lim_{\ell\to\infty} M^\ell_{i,j} = m_j$. This argument is pushed further in Section 1.2.3 below.

PROOF (2) OF THEOREM 1.2.1: For any square matrix let

$$\rho(A) = \max\{|\lambda| : \lambda \text{ an eigenvalue of } A\}.$$

Observe that any norm $\|\cdot\|$ on matrices that is submultiplicative (i.e., $\|AB\| \leq \|A\|\|B\|$) must satisfy $\rho(A) \leq \|A\|$.

Lemma 1.2.3 *For any square matrix A and any $\epsilon > 0$ there exists a submultiplicative matrix norm $\|\cdot\|$ such that $\|A\| \leq \rho(A) + \epsilon$.*

PROOF: Let U be a unitary matrix such that $A' = UAU^*$ with A' upper-triangular. Let $D = D(t)$, $t > 0$, be the diagonal matrix with $D_{i,i} = t^i$. Then $A'' = DA'D^{-1}$ is upper-triangular with $A''_{i,j} = t^{-(j-i)}A'_{i,j}$, $j \geq i$. Note that, by construction, the diagonal entries are the eigenvalues of A. Consider the matrix norm (induced by the vector norm $\|v\|_1 = \sum |v_i|$)

$$\|B\|_1 = \max_j \sum_i |B_{i,j}|.$$

Then $\|A''\|_1 = \rho(A) + O(t^{-1})$. Pick $t > 0$ large enough so that $\|A''\|_1 \leq \rho + \epsilon$. For U, D fixed as above, define a matrix norm by setting, for any matrix B,

$$\|B\| = \|DUBU^*D^{-1}\|_1 = \|(UD)B(DU)^{-1})\|_1.$$

This norm satisfies the conclusion of the lemma (observe that it depends very much on A and ϵ).

Lemma 1.2.4 *We have* $\lim_{\ell\to\infty} \max_{i,j} A^\ell_{i,j} = 0$ *if and only if* $\rho(A) < 1$.

For each $\epsilon > 0$, the submultiplicative norm of Lemma 1.2.3 satisfies

$$\|A\| \leq \rho(A) + \epsilon.$$

If $\rho(A) < 1$, then we can pick $\epsilon > 0$ so that $\|A\| < 1$. Then $\lim_{\ell\to\infty} \|A^\ell\| \leq \lim_{\ell\to\infty} \|A\|^k = 0$. The desired conclusion follows from the fact that all norms on a finite dimensional vector space are equivalent. Conversely, if

$$\lim_{\ell\to\infty} \left(\max_{i,j} A^\ell_{i,j} \right) = 0$$

then $\lim_{\ell\to\infty} \|A^\ell\|_1 = 0$. Since $\|\cdot\|_1$ is multiplicative, $\rho(A) \leq \|A^\ell\|_1^{1/\ell} < 1$ for ℓ large enough.

Let us pause here to see how the above argument translates in quantitative terms. Let $\|A\|_\infty = \max_{i,j} |A_{i,j}|$ and $\|\!|A|\!\|^2 = \sum_{i,j} |A_{i,j}|^2$. We want to bound $\|A^\ell\|_\infty$ in terms of the norm $\|A^\ell\|$ of Lemma 1.2.3.

Lemma 1.2.5 *For any $n \times n$ matrix A and any $\epsilon > 0$, we can choose the norm $\|\cdot\|$ of Lemma* 1.2.3 *so that*

$$\|A^\ell\|_\infty \leq n^{1/2}(1 + \|\!|A|\!\|/\epsilon)^n \|A^\ell\|.$$

PROOF: With the notation of the proof of Lemma 1.2.3, we have

$$\begin{aligned} |A'_{i,j}| &= \sum_{s,t} U_{i,s} A_{s,t} \overline{U}_{j,t} \\ &\leq \left(\sum_{s,t} |A_{s,t}|^2 \right)^{1/2} \left(\sum_{s,t} |U_{i,s}|^2 |U_{j,t}|^2 \right)^{1/2} \\ &\leq \left(\sum_{s,t} |A_{s,t}|^2 \right)^{1/2} \leq \|\!|A|\!\| \end{aligned}$$

because U is unitary. It follows that

$$\sum_i |A''_{i,j}| \leq \rho(A) + |||A|||(t-1)^{-1}.$$

Hence, for $t = 1 + |||A|||/\epsilon$, we get

$$\|A\| = \|A''\|_1 \leq \rho(A) + \epsilon$$

as desired. Now, for any ℓ, set $B = A^\ell, B' = (A')^\ell, B'' = (A'')^\ell$. Then $\|A^\ell\| = \|B''\|_1$ and $A^\ell = U^*B'U = U^*D^{-1}B''DU$. The matrix $B' = D^{-1}B''D$ is upper-triangular with coefficients $B'_{i,j} = t^{j-i}B''_{i,j}$ for $j \geq i$. This yields

$$\begin{aligned}
\|A^\ell\|_\infty &\leq \left(\sum_{\substack{i,j:\\ i\leq j}} t^{2(j-i)}|B''_{i,j}|^2\right)^{1/2} \\
&\leq n^{1/2}(1+|||A|||/\epsilon)^n\|B''\|_1 \\
&= n^{1/2}(1+|||A|||/\epsilon)^n\|A^\ell\|.
\end{aligned}$$

With this material at hand the following lemma suffices to finish the second proof or the Perron-Frobenius theorem.

Lemma 1.2.6 *Let M be a stochastic matrix satisfying the strong irreducibility condition of Theorem* 1.2.1. *Let $M^\infty_{i,j} = m_j$ where $m = (m_j)$ is the unique normalized row vector with positive entries such that $mM = m$. Then $\rho(M - M^\infty) < 1$.*

PROOF: Let λ be an eigenvalue of M with left eigenvector v. Assume that $|\lambda| = 1$. Then, again, $|v|$ is a left eigenvector with eigenvalue 1. Let k be such that $M^k > 0$. It follows that

$$|\sum_j M^k_{i,j}v_j| = \sum_j M^k_{i,j}|v_j|.$$

Since $M^k_{i,j} > 0$ for all j, this implies that $v_j = e^{i\theta}|v_j|$ for some fixed θ. Hence $\lambda = 1$. Let $\lambda_1 = 1$ and $\lambda_i,\ i = 2,\dots,n$ be the eigenvalues of M repeated according to there geometric multiplicities. By Lemma 1.2.2, $|\lambda_i| < 1$ for $i = 2,\dots,n$. The eigenvalues of M^∞ are 1 with eigenspace $\mathbb{R}\mathbf{1}$ and 0 with eigenspace $V_0 = \{v : \sum_i v_i = 0\}$. By (1.2.2) it follows that the eigenvalues of $M - M^\infty$ are $0 = \lambda_1 - 1$ and $\lambda_i = \lambda_i - 0$, $i = 2,\dots,n$. Hence $\rho(M - M^\infty) < 1$.

1.2.2 Comments on the Perron-Frobenius theorem

Each of the two proofs of Theorem 1.2.1 outlined above provides existence of $A > 0$ and $0 < \epsilon < 1$ such that

$$|M^\ell_{i,j} - m_j| \leq A(1-\epsilon)^\ell. \tag{1.2.3}$$

However, it is rather dishonest to state the conclusion (1.2.1) in this form without a clear WARNING:

$$\boxed{\text{the proof does not give a clue on how large } A \text{ and how small } \epsilon \text{ can be.}}$$

Indeed, "Proof (1)" looks like a quantitative proof since it shows that

$$|M_{i,j}^{\ell} - m_j| \leq (1-c)^{\lfloor \ell/k \rfloor} \tag{1.2.4}$$

whenever $M^k \geq cM^\infty$. But, in general, it is hard to find explicit reasonable k and c such that the condition $M^k \geq cM^\infty$ is satisfied.

EXAMPLE 1.2.1: Consider the random walk on $\mathbb{Z}/n\mathbb{Z}$, $n = 2p+1$, where, at each step, we add 1 or substract 1 or do nothing each with probability 1/3. Then M is an $n \times n$ matrix with $M_{i,j} = 1/3$ if $|i-j| = 0, 1$, $M_{1,n} = M_{n,1} = 1/3$, and all the orther entries equal to zero. The matrix M^∞ has all its entries equal to $1/n$. Obviously, $M^p \geq n\,3^{-p}\,M^\infty$, hence $|M_{i,j}^{\ell} - (1/n)| \leq 2(1 - n3^{-p})^{\lfloor \ell/p \rfloor}$. This is a very poor estimate. It is quite typical of what can be obtained by using (1.2.4).

Still, there is an interesting conclusion to be drawn from (1.2.4). Let

$$k_0 = \inf\{\ell : M^\ell \geq (1 - 1/e)M^\infty\}$$

where the constant $c = 1 - 1/e$ as been chosen for convenience. This k_0 can be interpreted as a measure of how long is takes for the chain to be close to equilibrium in a crude sense. Then (1.2.4) says that this crude estimate suffices to obtain the exponential decay with rate $1/k_0$

$$|M_{i,j}^{\ell} - m_j| \leq 3e^{-\ell/k_0}$$

"Proof (2)" has the important theoretical advantage of indicating what is the best exponential rate in (1.2.3). Namely, for any norm $\|\cdot\|$ on matrices, we have

$$\lim \|M^\ell - M^\infty\|^{1/\ell} = \rho \tag{1.2.5}$$

where

$$\rho = \rho(M - M^\infty) = \max\{|\lambda| : \lambda \neq 1, \lambda \text{ an eigenvalue of } M\}.$$

Comparing with (1.2.4) we discover that $M^k \geq cM^\infty$ implies

$$\rho \leq \frac{1}{k}\log(1-c).$$

Of course (1.2.5) shows that, for all $\epsilon > 0$, there exists $C(\epsilon)$ such that

$$|M_{i,j}^{\ell} - m_j| \leq C(\epsilon)\,(\rho + \epsilon)^\ell.$$

The constant $C(\epsilon)$ can be large and is dificult to bound. Since $\|\!|M^\ell - M^\infty|\!\| \le 2n^{1/2}$ (in the notation of the proof of Lemma 1.2.5), Lemma 1.2.5 yields

$$|M^\ell_{i,j} - m_j| \le n^{1/2} \left(1 + \frac{2n^{1/2}}{\epsilon}\right)^n (\rho+\epsilon)^\ell. \tag{1.2.6}$$

This is quantitative, but essentially useless. I am not sure what is the best possible universal estimate of this sort but I find the next example quite convincing in showing that "Proof (2)" is not satisfactory from a quantitative point of view.

EXAMPLE 1.2.2: Let $\mathcal{X} = \{0,1\}^n$. Define a Markov chain with state space $\mathcal{X}$ as follows. If the current state is $x = (x_1,\dots,x_n)$ then move to $y = (y_1,\dots,y_n)$ where $y_i = x_{i+1}$ for $i = 1,\dots,n-1$ and $y_n = x_1$ or $y_n = x_1 + 1$ (mod 2), each with equal probability 1/2. It is not hard to verify that this chain is irreducible. Let M denote the matrix of this chain for some ordering of the state space. Then the left normalized eigenvector m with eigenvalue 1 is the constant vector with $m_i = 2^{-n}$. Furthermore, a moment of thought shows that $M^n = M^\infty$. Hence $\rho = \rho(M - M^\infty) = 0$. Now, $\max_{i,j} |M^{n-1}_{i,j} - m_j|$ is of order 2^{-n}. So, in this case, $C(\varepsilon)$ of order $(2\epsilon)^{-n}$ is certainly needed for the inequality $|M^\ell_{i,j} - m_j| \le C(\epsilon)\,(\rho+\epsilon)^\ell$ to be satisfied for all ℓ.

1.2.3 Further remarks on strong irreducibility

A n-dimensional stochastic matrix M is strongly irreducible if there exists an integer k such that, for all i,j, $M^k_{i,j} > 0$. This is related to what is known as the **Doeblin** condition. Say that M satisfies the Doeblin condition if there exist an integer k, a positive c, and a probability measure q on $\{1,\dots,n\}$ such that

$$\text{(D)} \qquad \text{for all } i \in \{1,\dots,n\}, \quad M^k_{i,j} \ge c q_j.$$

Proof (1) of Theorem 1.2.1 is based on the fact that strong irreducibility implies the Doeblin condition (**D**) with $q = m$ (the stationary measure) and some $k, c > 0$. The argument developed in this case yields the following well known result.

Theorem 1.2.7 *If M satisfies (**D**) for some $k, c > 0$ and a some probability q then*

$$\sum_j |M^\ell_{i,j} - m_j| \le 2(1-c)^{\lfloor \ell/k \rfloor}$$

for all integer ℓ. Here $m = (m_j)_1^n$ is the vector appearing in Lemma 1.2.2, i.e., the stationary measure of M.

PROOF: Using (1.2.1), observe that (**D**) implies $m_j \ge c q_j$. Let M^∞ be the matrix with all rows equal to m, let Q be the matrix with all rows equal to q and set

$$N = \frac{1}{1-c}\left(M^k - cQ\right), \quad N^\infty = \frac{1}{1-c}\left(M^\infty - cQ\right).$$

These two matrices are stochatic. Furthermore

$$M^k - M^\infty = (1-c)\,(N - N^\infty)$$

and

$$\begin{aligned} M^{k\ell} - M^\infty &= \left(M^k - M^\infty\right)^\ell \\ &= (1-c)^\ell\,(N - N^\infty)^\ell\,. \end{aligned}$$

Observe that $(N - N^\infty)^2 = (N - N^\infty)N$ because N^∞ has constant columns so that $PN^\infty = N^\infty$ for any stochastic matrix P. It follows that $(N - N^\infty)^\ell = (N - N^\infty)N^{\ell-1}$. If we set $\|\!|A\|\!|_1 = \max_i \sum_j |A_{i,j}|$ for any matrix A and recall that $\|\!|AB\|\!|_1 \le \|\!|A\|\!|_1 \|\!|B\|\!|_1$ we get

$$\|\!|M^{k\ell} - M^\infty\|\!|_1 \le (1-c)^\ell \|\!|N - N^\infty\|\!|_1 \|\!|N^{\ell-1}\|\!|_1.$$

Since N is stochastic, we have $\|\!|N\|\!|_1 = 1$. Also $\|\!|N - N^\infty\|\!|_1 \le 2$. Hence

$$\max_i \sum_j |M^{k\ell} - M^\infty| \le 2(1-c)^\ell.$$

This implies the stated result because $\ell \to \|\!|M^\ell - M^\infty\|\!|_1$ is nonincreasing.

1.3 Elementary functional analysis

This section introduces notation and concepts from elementary functional analysis such as operator norms, interpolation, and duality. This tools turn out to be extremely useful in manipulating finite Markov chains.

1.3.1 Operator norms

Let A, B be two Banach spaces with norms $\|\cdot\|_A,\ \|\cdot\|_B$. Let $K : A \to B$ be a linear operator. We set

$$\|K\|_{A\to B} = \sup_{\substack{f\in A:\\ \|f\|_A\le 1}} \{\|Kf\|_B\} = \sup_{f\in A: f\neq 0} \left\{\frac{\|Kf\|_B}{\|f\|_A}\right\}.$$

If A^*, B^* are the (topological) duals of A, B, the dual operator $K^* : B^* \to A^*$ defined by $K^* b^*(a) = b^*(Ka), a \in A$, satisfies

$$\|K^*\|_{B^*\to A^*} \le \|K\|_{A\to B}.$$

In particular, if $\mathcal{X}$ is a countable set equipped with a positive measure π and if $A = \ell^p(\pi)$ and $B = \ell^q(\pi)$ with

$$\|f\|_p = \|f\|_{\ell^p(\pi)} = \left(\sum_{x\in\mathcal{X}} |f(x)|^p \pi(x)\right)^{1/p} \quad \text{and} \quad \|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)|,$$

we write

$$\|K\|_{p\to q} = \|K\|_{\ell^p(\pi)\to\ell^q(\pi)}.$$

Let

$$\langle f,g\rangle = \langle f,g\rangle_\pi = \sum_{\mathcal{X}} f(x)\overline{g(x)}\pi(x)$$

be the scalar product on $\ell^2(\pi)$. For $1\le p<\infty$, this scalar product can be used to identify $\ell^p(\pi)^*$ with $\ell^q(\pi)$ where p,q are Hölder conjugate exponents, that is $1/p+1/q=1$. Furthermore, for all $1\le p\le\infty$, $\ell^q(\pi)$ norms $\ell^p(\pi)$. Namely,

$$\|f\|_p = \sup_{\substack{g\in\ell^q(\pi)\\ \|g\|_q\le 1}} \langle f,g\rangle_\pi.$$

It follows that for any linear operator $K:\ell^p(\pi)\to\ell^r(\pi)$ with $1\le p,r\le+\infty$,

$$\|K\|_{p\to r} = \|K^*\|_{s\to q}$$

where $1/p+1/q=1$, $1/r+1/s=1$. Assume now that the operator K is defined by

$$Kf(x) = \sum_{y\in\mathcal{X}} K(x,y)f(y)$$

for any finitely supported function f. Then the norm $\|K\|_{p\to\infty}$ is given by

$$\|K\|_{p\to\infty} = \max_{x\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}} |K(x,y)/\pi(y)|^q\pi(y)\right)^{1/q} \tag{1.3.1}$$

where $1/p+1/q=1$. In particular,

$$\|K\|_{2\to\infty} = \|K^*\|_{1\to 2} = \max_{x\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}} |K(x,y)/\pi(y)|^2\pi(y)\right)^{1/2} \tag{1.3.2}$$

and

$$\|K\|_{1\to\infty} = \|K^*\|_{1\to\infty} = \max_{x,y\in\mathcal{X}}\{|K(x,y)/\pi(y)|\}. \tag{1.3.3}$$

For future reference we now recall the Riesz-Thorin interpolation theorem (complex method). It is a basic tools in modern analysis. See, e.g., Theorem 1.3, page 179 in [73].

Theorem 1.3.1 *Fix $1\le p_i,q_i\le\infty$, $i=1,2$, with $p_1\le p_2$, $q_1\le q_2$. Let K be a linear operator acting on functions by $Kf(x)=\sum_y K(x,y)f(y)$. For any p such that $p_1\le p\le p_2$ let θ be such that $1/p=\theta/p_1+(1-\theta)/p_2$ and define $q\in[q_1,q_2]$ by $1/q=\theta/q_1+(1-\theta)/q_2$. Then*

$$\|K\|_{p\to q} \le \|K\|_{p_1\to q_1}^{\theta}\|K\|_{p_2\to q_2}^{1-\theta}.$$

1.3.2 Hilbert space techniques

For simplicity we assume now that $\mathcal{X}$ is finite of cardinality $n = |\mathcal{X}|$ and work on the (n-dimensional) Hilbert space $\ell^2(\pi)$. An operator $K : \ell^2(\pi) \to \ell^2(\pi)$ is self-adjoint if it satisfies

$$\langle Kf, g\rangle_\pi = \langle f, Kg\rangle_\pi, \quad \text{i.e.,} \quad K^* = K.$$

Let $K(x,y)$ be the kernel of the operator K. Then K^* has kernel

$$K^*(x,y) = \pi(y)\overline{K(y,x)}/\pi(x)$$

and it follows that K is selfadjoint if and only if

$$K(x,y) = \pi(y)\overline{K(y,x)}/\pi(x).$$

Lemma 1.3.2 *Assume that K is self-adjoint on $\ell^2(\pi)$. Then K is diagonalizable in an orthonormal basis of $\ell^2(\pi)$ and has real eigenvalues $\beta_0 \geq \beta_1 \ldots \geq \beta_{n-1}$. For any associated orthonormal basis $(\psi_i)_0^{n-1}$ of eigenfunctions, we have*

$$K(x,y)/\pi(y) = \sum_i \beta_i \psi_i(x)\overline{\psi_i(y)}. \tag{1.3.4}$$

$$\|K(x,\cdot)/\pi(\cdot)\|_2^2 = \sum_i \beta_i^2 |\psi_i(x)|^2. \tag{1.3.5}$$

$$\sum_{x\in\mathcal{X}} \|K(x,\cdot)/\pi(\cdot)\|_2^2 \pi(x) = \sum_i \beta_i^2. \tag{1.3.6}$$

PROOF: We only prove the set of equalities. Let $z \to \mathbf{1}_x(z)$ be the function which is equal to 1 at x and zero everywhere else. Then $K(x,y) = K\mathbf{1}_y(x)$. The function $\mathbf{1}_y$ has coordinates $\langle \mathbf{1}_y, \psi_i\rangle_\pi = \overline{\psi_i(y)}\pi(y)$ in the orthonormal basis $(\psi_i)_0^{n-1}$. Hence $K\mathbf{1}_y(x) = \pi(y)\sum_i \beta_i\psi_i(x)\overline{\psi_i(y)}$. The second and third results follow by using the fact that $(\psi_i)_0^{n-1}$ is orthonormal.

We now turn to an important tool known as the Courant-Fischer min-max theorem. Let $\mathcal{E}$ be a (positive) Hermitian form on $\ell^2(\pi)$. For any vector space $W \subset \ell^2(\pi)$, set

$$M(W) = \max_{\substack{f\in W\\ f\neq 0}} \left\{ \frac{\mathcal{E}(f,f)}{\|f\|_2^2} \right\}, \quad m(W) = \min_{f\in W} \left\{ \frac{\mathcal{E}(f,f)}{\|f\|_2^2} \right\}.$$

Recall from linear algebra that there exists a unique Hermitian matrice A such that $\mathcal{E}(f,f) = \langle Af, f\rangle_\pi$ and that, by definition, the eigenvalues of $\mathcal{E}$ are the eigenvalues of A. Furthermore, these are real.

Theorem 1.3.3 *Let $\mathcal{E}$ be a quadratic form on $\ell^2(\pi)$, with eigenvalues*

$$\lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{n-1}.$$

Then

$$\lambda_k = \min_{\substack{W\subset\ell^2(\pi):\\ \dim(W)\geq k+1}} M(W) = \max_{\substack{W\subset\ell^2(\pi):\\ \dim(W^\perp)\leq k}} m(W). \tag{1.3.7}$$

For a proof, see [51], page 179-180. Clearly, the minimum of $M(W)$ with $\dim(W) \geq k+1$ is obtained when W is the linear space spanned by the $k+1$ first eigenvectors ψ_i associated with λ_i, $i = 0, \dots, k$. Similarly, the maximum of $m(W)$ with $\dim(W^\perp) \leq k$ is attained when W is spanned by the ψ_i's, $i = k, \dots, n$. This result also holds in infinite dimension. It has the following corollary.

Theorem 1.3.4 *Let $\mathcal{E}, \mathcal{E}'$ be two quadratic forms on different Hilbert spaces $\mathcal{H}$, $\mathcal{H}'$ of dimension $n \leq n'$. Assume that there exists a linear map $f \to \tilde{f}$ from $\mathcal{H}$ into $\mathcal{H}'$ such that, for all $f \in \mathcal{H}$,*

$$\mathcal{E}'(\tilde{f}, \tilde{f}) \leq A\mathcal{E}(f, f) \quad \textit{and} \quad a\|f\|_{\mathcal{H}}^2 \leq \|\tilde{f}\|_{\mathcal{H}'}^2 \tag{1.3.8}$$

for some constants $0 < a, A < \infty$. Then

$$\frac{a}{A}\,\lambda'_\ell \leq \lambda_\ell \quad \textit{for} \quad \ell = 1, \dots, n-1. \tag{1.3.9}$$

PROOF: Fix $\ell = 0, 1, \dots, n-1$ and let ψ_i be orthonormal eigenvectors associated to λ_i, $i = 0, \dots, n-1$. Observe that the second condition in (1.3.8) implies that $f \to \tilde{f}$ is one to one. Let $W \subset \mathcal{H}$ be the vector space spanned by $(\psi_i)_0^{\ell-1}$, and let $\widetilde{W} \subset \mathcal{H}'$ be its image under the one to one map $f \to \tilde{f}$. Then $\widetilde{W}$ has dimension ℓ and by (3.7)

$$\begin{aligned} \lambda'_\ell &\leq M(\widetilde{W}) = \max_{f \in W} \left\{ \frac{\mathcal{E}'(\tilde{f}, \tilde{f})}{\|\tilde{f}\|_{\mathcal{H}'}^2} \right\} \\ &\leq \max_{f \in W} \left\{ \frac{A\mathcal{E}(f, f)}{a\|f\|_{\mathcal{H}}^2} \right\} = \frac{A\lambda_\ell}{a}. \end{aligned}$$

1.4 Notation for finite Markov chains

Let $\mathcal{X}$ be a finite space of cardinality $|\mathcal{X}| = n$. Let $K(x, y)$ be a Markov kernel on $\mathcal{X}$ with associated Markov operator defined by

$$Kf(x) = \sum_{y \in \mathcal{X}} K(x, y) f(y).$$

That is, we assume that

$$K(x, y) \geq 0 \quad \text{and} \quad \sum_y K(x, y) = 1.$$

The operator K^ℓ has a kernel $K^\ell(x, y)$ which satisfies

$$K^\ell(x, y) = \sum_{z \in \mathcal{X}} K^{\ell-1}(x, z) K(z, y).$$

Properly speaking, the Markov chain with initial distribution q associated with K is the sequence of $\mathcal{X}$-valued random variables $(X_n)_0^\infty$ whose law $\mathbf{P}_q$ is determined by

$$\forall\, \ell = 1, 2, \dots, \quad \mathbf{P}_q(X_i = x_i, 1 \le i \le \ell) = q(x_0)K(x_0, x_1)\cdots K(x_{\ell-1}, x_\ell).$$

With this notation the probability measure $K^\ell(x, \cdot)$ is the law of X_ℓ for the Markov chain started at x:

$$\mathbf{P}_x(X_\ell = y) = K^\ell(x, y).$$

However, this language will almost never be used in these notes.

The continuous time semigroup associated with K is defined by

$$H_t f(x) = e^{-t(I-K)} = e^{-t}\sum_0^\infty \frac{t^i K^i f}{i!}. \tag{1.4.1}$$

Obviously, it has kernel

$$H_t(x, y) = e^{-t}\sum_0^\infty \frac{t^i K^i(x, y)}{i!}.$$

Observe that this is indeed a semigroup of operators, that is,

$$\begin{aligned} H_{t+s} &= H_t H_s \\ \lim_{t \to 0} H_t &= I. \end{aligned}$$

Furthermore, for any f, the function $u(t, x) = H_t f(x)$ solves

$$\begin{cases} (\partial_t + (I - K))\, u(t, x) = 0 \text{ on } (0, \infty) \times \mathcal{X} \\ u(0, x) = f(x). \end{cases}$$

Set $H_t^x(y) = H_t(x, y)$. Then $H_t^x(\cdot)$ is a probability measure on $\mathcal{X}$ which represents the distribution a time t of the continuous Markov chain $(X_t)_{t>0}$ associated with K and started at x. This process can be described as follows. The moves are those of the discrete time Markov chain with transition kernel K started at x, but the jumps occur after independent Poison(1) waiting times. Thus, the probability that there have been exactly i jumps at time t is $e^{-t}\, t^i/i!$ and the probability to be at y after exactly i jumps at time t is $e^{-t}\, t^i\, K^i(x, y)/i!$.

The operators K, H_t also acts on measures. If μ is a measure then μK (resp. μH_t) is defined by setting

$$\mu K(f) = \mu(Kf) \quad (\text{resp. } \mu H_t(f) = \mu(H_t f))$$

for all functions f. Thus

$$\mu K(x) = \sum_y \mu(y) K(y, x).$$

Definition 1.4.1 *A Markov kernel K on a finite set $\mathcal{X}$ is said to be irreducible if for any x, y there exists $j = j(x, y)$ such that $K^j(x, y) > 0$.*

Assume that K is irreducible and let π be the unique stationary measure for K, that is, the unique probability measure satisfying $\pi K = \pi$ (see Lemma 1.2.2). We will use the notation

$$\pi(f) = \sum_x f(x)\pi(x) \quad \text{and} \quad \mathrm{Var}_\pi(f) = \sum_x |f(x) - \pi(f)|^2\pi(x).$$

We also set

$$\pi_* = \min_{x\in\mathcal{X}}\{\pi(x)\}. \tag{1.4.2}$$

Throughout these notes we will work with the Hilbert space $\ell^2(\pi)$ with scalar product

$$\langle f, g\rangle = \sum_{x\in\mathcal{X}} f(x)\overline{g(x)}\pi(x),$$

and with the space $\ell^p(\pi)$, $1 \le p \le \infty$, with norm

$$\|f\|_p = \left(\sum_{x\in\mathcal{X}} |f(x)|^p\pi(x)\right)^{1/p}, \quad \|f\|_\infty = \max_{x\in\mathcal{X}}\{|f(x)|\}.$$

In this context, it is natural and useful to consider the densities of the probability measures K_x^ℓ, H_t^x with respect to π which will be denoted by

$$k_x^\ell(y) = k^\ell(x, y) = \frac{K^\ell(x, y)}{\pi(y)}$$

and

$$h_t^x(y) = h_t(x, y) = \frac{H_t^x(y)}{\pi(y)}.$$

Observe that the semigroup property implies that, for all $t, s > 0$,

$$h_{t+s}(x, y) = \sum_z h_t(x, z)h_s(z, y)\pi(z).$$

The operator K (hence also H_t) is a contraction on each $\ell^p(\pi)$ (i.e., $\|Kf\|_p \le \|f\|_p$). Indeed, by Jensen's inequality, $|Kf(x)|^p \le K(|f|^p)(x)$ and thus

$$\|Kf\|_p^p \le \sum_{x,y} K(x, y)|f(y)|^p\pi(x) = \sum_y |f(y)|^p\pi(y) = \|f\|_p^p.$$

The adjoint K^* of K on $\ell^2(\pi)$ has kernel

$$K^*(x, y) = \pi(y)K(y, x)/\pi(x).$$

Since π is the stationary measure of K, it follows that K^* is a Markov operator. The associated semigroup is $H_t^* = e^{-t(I-K^*)}$ with kernel

$$H_t^*(x, y) = \pi(y)H_t(y, x)/\pi(x)$$

and density

$$h_t^*(x,y) = h_t(y,x).$$

The Markov process associated with H_t^* is the time reversal of the process associated to H_t.

If a measure μ has density f with respect to π, that is, if $\mu(x) = f(x)\pi(x)$, then μK (resp. μH_t) has density $K^* f$ (resp. $H_t^* f$) with respect to π. Thus acting by K (resp. H_t) on a measure is equivalent to acting by K^* (resp H_t^*) on its density with respect to π. In particular, the density $h_t(x,\cdot)$ of the measure H_t^x with respect to π is $H_t^*\delta_x$ where $\delta_x = \mathbf{1}_x/\pi(x)$. Indeed, the measure $\mathbf{1}_x$ has density $\delta_x = \mathbf{1}_x/\pi(x)$ with respect to π. Hence $H_t^x = \mathbf{1}_x H_t$ has density

$$H_t^*\delta_x(y) = \frac{H_t^*(y,x)}{\pi(x)} = h_t^*(y,x) = h_t(x,y)$$

with respect to π.

Recall the following classic definition.

Definition 1.4.2 *A pair (K,π) where K is Markov kernel and π a positive probability measure on $\mathcal{X}$ is reversible if*

$$\pi(x)K(x,y) = \pi(y)K(y,x).$$

This is sometimes called the detailed balance condition.

If (K,π) is reversible then $\pi K = \pi$. Furthermore, (K,π) is reversible if and only if K is self-adjoint on $\ell^2(\pi)$.

1.4.1 Discrete time versus continuous time

These notes are written for continuous time finite Markov chains. The reason of this choice is that it makes life easier from a technical point of view. This will allow us hopefully to stay more focussed on the main ideas. This choice however is not very satisfactory because in some respects (e.g., implementation of algorithms) discrete time chains are more natural. Furthermore, since the continuous time chain is obtained as a function of the discrete time chain through the formula $H_t = e^{-t(I-K)}$ it is often straightforward to transfer information from discrete time to continuous time whereas the converse can be more difficult. Thus, let us emphasize that the techniques presented in these lectures are not confined to continuous time and work well in discrete time. Treatments of discrete time chains in the spirit of these notes can be found in [23, 24, 25, 26, 27, 28, 29, 35, 41, 63].

For reversible chains, it is possible to relate precisely the behavior of H_t to that of K^ℓ through eigenvalues and eigenvectors as follows. Assuming that (K,π) is reversible and $|\mathcal{X}| = n$, let $(\lambda_i)_0^{n-1}$ be the eigenvalues of $I-K$ in non-decreasing order and let $(\psi_i)_0^{n-1}$ be an orthonormal basis of $\ell^2(\pi)$ made of real eigenfuntions associated to the eigenvalues $(\lambda_i)_0^{n-1}$ with $\psi_0 \equiv 1$.

Lemma 1.4.3 *If (K,π) is reversible, it satisfies*

(1) $k^\ell(x,y) = \sum_0^{n-1}(1-\lambda_i)^\ell \psi_i(x)\psi_i(y), \quad \|k_x^\ell - 1\|_2^2 = \sum_1^{n-1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2.$

(2) $h_t(x,y) = \sum_0^{n-1} e^{-t\lambda_i}\psi_i(x)\psi_i(y), \ \|h_t^x - 1\|_2^2 = \sum_1^{n-1} e^{-2t\lambda_i}|\psi_i(x)|^2.$

This classic result follows from Lemma 1.3.2. The next corollary gives a useful way of transferring information between discrete and continuous time. It separates the effects of the largest eigenvalue λ_{n-1} from those of the rest of the spectrum.

Corollary 1.4.4 *Assume that (K,π) is reversible and set $\beta_- = \max\{0, -1 + \lambda_{n-1}\}$. Then*

(1) $\|h_t^x - 1\|_2^2 \leq \frac{1}{\pi(x)}e^{-t} + \|k_x^{[t/2]} - 1\|_2^2.$

(2) $\|k_x^N - 1\|_2^2 \leq \beta_-^{2m}\left(1 + \|h_\ell^x - 1\|_2^2\right) + \|h_N^x - 1\|_2^2$ *for* $N = m + \ell + 1$.

Proof: For (1), use Lemma 1.4.3,

$$(1-\lambda_i)^{2\ell} = e^{2\ell\log(1-\lambda_i)}$$

and the inequality $\log(1-x) \geq -2x$ for $0 \leq x \leq 1/2$. For (2), observe that

$$k^{2\ell+1}(x,x) = \sum_0^{n-1}(1-\lambda_i)^{2\ell+1}|\psi_i(x)|^2 \geq 0.$$

This shows that

$$-\sum_{i:\lambda_i>1}(1-\lambda_i)^{2\ell+1}|\psi_i(x)|^2 \leq \sum_{i:\lambda_i<1}(1-\lambda_i)^{2\ell+1}|\psi_i(x)|^2.$$

Hence

$$\sum_{i:\lambda_i>1}(1-\lambda_i)^{2\ell+2}|\psi_i(x)|^2 \leq \sum_{i:\lambda_i<1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2.$$

Now, for those λ_i that are smaller than 1, we have

$$(1-\lambda_i)^{2\ell} = e^{2\ell\log(1-\lambda_i)} \leq e^{-2\ell\lambda_i}$$

so that

$$\sum_{i:\lambda_i<1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2 \leq \|h_\ell^x\|_2^2$$

and

$$\sum_{i\neq 0,\lambda_i<1}(1-\lambda_i)^{2\ell}|\psi_i(x)|^2 \leq \|h_\ell^x - 1\|_2^2.$$

Putting these pieces together, we get for $N = m + \ell + 1$,

$$\begin{aligned}
\|k_x^N - 1\|_2^2 &= \sum_1^{n-1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 \\
&= \sum_{i:\lambda_i>1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 + \sum_{i\neq 0:\lambda_i<1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 \\
&\leq \beta_-^{2m} \left(\sum_{i:\lambda_i>1} (1-\lambda_i)^{2\ell+2} |\psi_i(x)|^2 \right) + \sum_{i\neq 0:\lambda_i<1} (1-\lambda_i)^{2N} |\psi_i(x)|^2 \\
&\leq \beta_-^{2m} \|h_\ell^x\|_2^2 + \|h_N^x - 1\|_2^2 \\
&= \beta_-^{2m} \left(1 + \|h_\ell^x - 1\|_2^2\right) + \|h_N^x - 1\|_2^2.
\end{aligned}$$

Observe that, according to Corrolary 1.4.4, it is useful to have tools to bound $1 - \lambda_{n-1}$ away from -1.

Corollary 1.4.4 says that the behavior of a discrete time chain and of its associated continuous time chain can not be too different in the reversible case. It is interesting to see that this fails to be satisfied for nonreversible chains.

EXAMPLE 1.4.1: Consider the chain K on $\mathcal{X} = \mathbb{Z}/m\mathbb{Z}$ with $m = n^2$ an odd integer and

$$K(x,y) = \begin{cases} 1/2 & \text{if } y = x+1 \\ 1/2 & \text{if } y = x+n \end{cases}.$$

On one hand, the discrete time chain takes order $m^2 \approx n^4$ steps to be close to stationarity. Indeed, there exists an affine bijection from $\mathcal{X}$ to $\mathcal{X}$ that send 1 to 1 and n to -1. On the other hand, one can show that the associated continuous time process is close to stationarity after a time of order $m = n^2$. See [25].

Lemma 1.4.3 is often hard to use directly because it involves both eigenvalues and eigenvectors. To have a similar statement involving only eigenvalues one has to work with the distance

$$|||f - g||| = \left(\sum_{x,y} |f(x,y) - g(x,y)|^2 \pi(x)\pi(y) \right)^{1/2}$$

between functions on $\mathcal{X} \times \mathcal{X}$.

Lemma 1.4.5 *If (K, π) is reversible, it satisfies*

$$|||k^\ell - 1|||^2 = \sum_1^{n-1} (1-\lambda_i)^{2\ell} \quad \text{and} \quad |||h_t - 1|||^2 = \sum_1^{n-1} e^{-2t\lambda_i}.$$

It is possible to bound $|||k^\ell - 1|||$ using only $\beta_* = \max\{1-\lambda_1, -1+\lambda_{n-1}\}$ and the eigenvalues λ_i such that $\lambda_i < 1$. It is natural to state this result in terms of the eigenvalues $\beta_i = 1 - \lambda_i$ of K. Then $\beta_* = \max\{\beta_1, |\beta_{n-1}|\}$ and $\lambda_i < 1$ corresponds to the condition $\beta_i > 0$.

Corollary 1.4.6 *Assume that (K,π) is reversible. With the notation introduced above we have, for $N = m + \ell + 1$,*

$$\|k^m - 1\|^2 \leq 2\beta_*^{2\ell} \left(\sum_{i:0<\beta_i\leq 1} \beta_i^{2m} \right).$$

PROOF: We have

$$\sum_{\mathcal{X}} k^{2m+1}(x,x)\pi(x) = \sum_0^{n-1} \beta_i^{2m+1} \geq 0.$$

Hence

$$\sum_{\beta_i<0} \beta_i^{2m+2} \leq \sum_{\beta_i>0} \beta_i^{2m}.$$

It follows that

$$\begin{aligned} \|k^N - 1\| &= \sum_1^{n-1} \beta_i^{2m+2\ell+2} \\ &\leq \beta_*^{2\ell} \left(\sum_0^{n-1} \beta_i^{2m+2} \right) \leq 2\beta_*^{2\ell} \left(\sum_{i:\beta_i>0} \beta_i^{2m} \right). \end{aligned}$$

Chapter 2

Analytic tools

This chapter uses semigroup techniques to obtain quantitative estimates on the convergence of continuous time finite Markov chain in terms of various functional inequalities. The same ideas and techniques apply to discrete time but the details are somewhat more tedious. See [28, 29, 35, 41, 63, 72].

2.1 Nothing but the spectral gap

2.1.1 The Dirichlet form

Classicaly, the notion of *Dirichlet form* is introduced in relation with reversible Markov semigroups. The next definition coincides with the classical notion when (K, π) is reversible.

Definition 2.1.1 *The form*

$$\mathcal{E}(f,g) = \Re(\langle (I-K)f, g\rangle)$$

is called the Dirichlet form associated with $H_t = e^{-t(I-K)}$

The notion of Dirichlet form will be one of our main technical tools.

Lemma 2.1.2 *The Dirichlet form* $\mathcal{E}$ *satisfies* $\mathcal{E}(f,f) = \langle (I - \frac{1}{2}(K+K^*))f, f\rangle$,

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y} |f(x)-f(y)|^2 K(x,y)\pi(x) \tag{2.1.1}$$

and

$$\frac{\partial}{\partial t}\|H_t f\|_2^2 = -2\,\mathcal{E}(H_t f, H_t f). \tag{2.1.2}$$

PROOF: The first equality follows from $\langle Kf, f\rangle = \langle f, K^* f\rangle = \overline{\langle K^* f, f\rangle}$. For the second, observe that $\mathcal{E}(f,f) = \|f\|_2^2 - \Re(\langle Kf, f\rangle)$ and

$$\frac{1}{2}\sum_{x,y} |f(x)-f(y)|^2 K(x,y)\pi(x)$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{x,y}\left(|f(x)|^2+|f(y)|^2-2\Re(\overline{f(x)}f(y))\right)K(x,y)\pi(x)\\
&= \|f\|_2^2-\Re(\langle Kf,f\rangle).
\end{aligned}
$$

The third is calculus. In a sense, (2.1.2) is the definition of $\mathcal{E}$ as the Dirichlet form of the semigroup H_t since

$$\mathcal{E}(f,f) = -\,\partial_t\|H_tf\|_2^2\big|_{t=0} = -\lim_{t\to 0}\frac{1}{t}\langle (I-H_t)f,f\rangle.$$

Lemma 2.1.2 shows that the Dirichlet forms of H_t, H_t^* and $S_t = e^{-t(I-R)}$ whith $R=\frac{1}{2}(K+K^*)$ are equal. Let us emphasize that equalities (2.1.1) and (2.1.2) are crucial in most developments involving Dirichlet forms. Equality (2.1.1) expresses the Dirichlet form as a sum of positive terms. It will allow us to estimate $\mathcal{E}$ in geometric terms and to compare different Dirichlet forms. Equality (2.1.2) is the key to translating functional inequalities such as Poincaré or logarithmic Sobolev inequalities into statements about the behavior of the semigroup H_t.

2.1.2 The spectral gap

This section introduces the notion of spectral gap and gives bounds on convergence that depend only on the spectral gap and the stationary measure.

Definition 2.1.3 *Let K be a Markov kernel with Dirichlet form $\mathcal{E}$. The spectral gap $\lambda=\lambda(K)$ is defined by*

$$\lambda = \min\left\{\frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)};\mathrm{Var}_\pi(f)\neq 0\right\}$$

Observe that λ is not, in general, an eigenvalue of $(I-K)$. If K is self-adjoint on $\ell^2(\pi)$ (that is, if (K,π) is reversible) then λ is the smallest non zero eigenvalue of $I-K$. In general λ is the smallest non zero eigenvalue of $I-\frac{1}{2}(K+K^*)$. Note also that the Dirichlet forms of K^* and K satisfy

$$\mathcal{E}_K(f,f)=\mathcal{E}_{K^*}(f,f).$$

It follows that $\lambda(K)=\lambda(K^*)$. Clearly, we also have

$$\lambda = \min\left\{\mathcal{E}(f,f);\|f\|_2=1,\ \pi(f)=0\right\}.$$

Furthermore, if one wishes, one can impose that f be real in the definition of λ. Indeed, let λ_r be the quantity obtained for real f. Then $\lambda_r\geq\lambda$ and, if $f=u+iv$ with u,v real functions, then $\lambda_r\mathrm{Var}_\pi(f)=\lambda_r(\mathrm{Var}_\pi(u)+\mathrm{Var}_\pi(v))\leq \mathcal{E}(v,v)+\mathcal{E}(u,u)=\mathcal{E}(f,f)$. Hence $\lambda_r\leq\lambda$ and finally $\lambda_r=\lambda$.

Lemma 2.1.4 *Let K be a Markov kernel with spectral gap $\lambda=\lambda(K)$. Then the semigroup $H_t=e^{-t(I-K)}$ satisfies*

$$\forall\ f\in\ell^2(\pi),\quad \|H_tf-\pi(f)\|_2^2\leq e^{-2\lambda t}\,\mathrm{Var}_\pi(f).$$

PROOF: Set $u(t) = \mathrm{Var}_\pi(H_t f) = \|H_t(f-\pi(f))\|_2^2 = \|H_t f - \pi(f)\|_2^2$. Then

$$u'(t) = -2\,\mathcal{E}\left(H_t(f-\pi(f)), H_t(f-\pi(f))\right) \le -2\lambda u(t).$$

It follows that

$$u(t) \le e^{-2\lambda\, t} u(0)$$

which is the desired inequality because $u(0) = \mathrm{Var}_\pi(f)$.

As a corollary we obtain one of the simplest and most useful quantitative results in finite Markov chain theory.

Corollary 2.1.5 *Let K be a Markov kernel with spectral gap $\lambda = \lambda(K)$. Then the density $h_t^x(\cdot) = H_t^x(\cdot)/\pi(\cdot)$ satisfies*

$$\|h_t^x - 1\|_2 \le \sqrt{1/\pi(x)}\; e^{-\lambda\, t}.$$

It follows that

$$|H_t(x,y) - \pi(y)| \le \sqrt{\pi(y)/\pi(x)}\; e^{-\lambda\, t}.$$

PROOF: Let H_t^* be the adjoint of H_t on $\ell^2(\pi)$ (see Section 2.1.1). This is a Markov semigroup with spectral gap $\lambda(K^*) = \lambda(K)$. Set $\delta_x(y) = 1/\pi(x)$ if $y = x$ and $\delta_x(y) = 0$ otherwise. Then

$$h_t^x(y) = \frac{H_t^x(y)}{\pi(y)} = H_t^* \delta_x(y)$$

and, by Lemma 2.1.4 applied to K^*,

$$\|H_t^*\delta_x - 1\|_2^2 \le e^{-2\lambda\, t}\mathrm{Var}_\pi(\delta_x).$$

Hence

$$\|h_t^x - 1\|_2 \le \sqrt{\frac{1-\pi(x)}{\pi(x)}}\; e^{-\lambda\, t} \le \frac{1}{\sqrt{\pi(x)}}\; e^{-\lambda\, t}.$$

Of course, the same result holds for H_t^*. Hence

$$\begin{aligned} |h_t(x,y) - 1| &= \left|\sum_z (h_{t/2}(x,z)-1)(h_{t/2}(z,y)-1)\pi(z)\right| \\ &\le \|h_{t/2}^x - 1\|_2 \|h_{t/2}^{*y} - 1\|_2 \\ &\le \frac{1}{\sqrt{\pi(x)\pi(y)}}\; e^{-\lambda\, t}. \end{aligned}$$

Multiplying by $\pi(y)$ yields the desired inequality. This ends the proof of Corollary 2.1.5.

Definition 2.1.6 *Let $\omega = \omega(K) = \min\{\Re(\zeta) : \zeta \neq 0$ an eigenvalue of $I - K\}$.*

Let $\mathcal{S}$ denote the spectrum of $I - K$. Since $H_t = e^{-t(I-K)}$, the spectrum of H_t is $\{e^{-t\xi} : \xi \in \mathcal{S}\}$. It follows that the spectral radius of $H_t - E_\pi$ in $\ell^2(\pi)$ is $e^{-t\omega}$. Using (1.2.5) we obtain the following result.

Theorem 2.1.7 *Let K be an irreducible Markov kernel. Then*

$$\forall\, 1 \le p \le \infty, \quad \lim_{t\to\infty} \frac{-1}{t} \log\left(\max_x \|h_t^x - 1\|_p\right) = \omega.$$

In particular, $\lambda \le \omega$ with equality if (K,π) is reversible. Furthermore, if we set

$$T_p = T_p(K, 1/e) = \min\left\{t > 0 : \max_x \|h_t^x - 1\|_p \le 1/e\right\}, \tag{2.1.3}$$

and define π_* as in (1.4.2) then, for $1 \le p \le 2$,

$$\frac{1}{\omega} \le T_p \le \frac{1}{2\lambda}\left(2 + \log\frac{1}{\pi_*}\right),$$

whereas, for for $2 < p \le \infty$,

$$\frac{1}{\omega} \le T_p \le \frac{1}{\lambda}\left(1 + \log\frac{1}{\pi_*}\right).$$

Example 2.1.1: Let $\mathcal{X} = \{0,\dots,n\}$. Consider the Kernel $K(x,y)) = 1/2$ if $y = x \pm 1$, $(x,y) = (0,0)$ or (n,n), and $K(x,y) = 0$ otherwise. This is a symmetric kernel with uniform stationary distribution $\pi \equiv 1/(n+1)$. Feller [40], page 436, gives the eigenvalues and eigenfunctions of K. For $I - K$, we get the following:

$$\lambda_0 = 0, \quad \psi_0(x) \equiv 1$$

$$\lambda_j = 1 - \cos\frac{\pi j}{n+1}, \quad \psi_j(x) = \sqrt{2}\cos(\pi j(x+1/2)/(n+1)) \quad \text{for} \quad j = 1,\dots,n.$$

Let $H_t = e^{-t(I-K)}$ and write (using $\cos(\pi x) \le 1 - 2x^2$ for $0 \le x \le 1$)

$$\begin{aligned} |h_t(x,y) - 1| &= \left|\sum_{j=1}^n \psi_j(x)\psi_j(y)e^{-t(1-\cos(\pi j/(n+1)))}\right| \\ &\le 2\sum_{j=1}^n e^{-2tj^2/(n+1)^2} \\ &\le 2e^{-2t/(n+1)^2}\left(1 + \sqrt{(n+1)^2/2t}\right). \end{aligned}$$

To obtain the last inequality, use

$$\sum_2^n e^{-2tj^2/(n+1)^2} \le \int_1^\infty e^{-2ts^2/(n+1)^2} ds = \frac{n+1}{\sqrt{2t}} \int_{\frac{\sqrt{2t}}{n+1}}^\infty e^{-u^2} du$$

and

$$\frac{2}{\sqrt{\pi}} \int_z^\infty e^{-u^2} du = \frac{2e^{-z^2}}{\sqrt{\pi}} \int_z^\infty e^{-(u-z)^2 - 2(u-z)z} du \le e^{-z^2}.$$

In particular,

$$\max_{x,y} |h_{2t}(x,y) - 1| = \max_x \|h_t^x - 1\|_2^2 \leq 2e^{-c} \quad \text{for} \quad t = \frac{1}{4}(n+1)^2(1+c)$$

and $T_2(K, 1/e) \leq 3(n+1)^2/4$. Also, $\omega = \lambda = 1 - \cos\frac{\pi}{n+1} \leq \pi^2/(n+1)^2$. Hence in this case, the lower bound for $T_2(K, 1/e)$ given by Theorem 2.1.6 is of the right order of magnitude whereas the upper bound

$$T_2 \leq \frac{1}{2\lambda}\left(2 + \log\frac{1}{\pi_*}\right) \leq \frac{1}{4}(n+1)^2(2 + \log(n+1))$$

is off by a factor of $\log(n+1)$.

EXAMPLE 2.1.2: Let $\mathcal{X} = \{0,1\}^n$ and $K(x,y) = 0$ unless $|x-y| = \sum_i |x_i - y_i| = 1$ in which case $K(x,y) = 1/n$. Viewing $\mathcal{X}$ as an Abelian group it is not hard to see that the characters

$$\chi_y : x \to (-1)^{y.x}, \quad y \in \{0,1\}^n$$

where $x.y = \sum_i x_i y_i$, form an orthonormal basis of $\ell^2(\pi)$, $\pi \equiv 2^{-n}$. Also

$$\begin{aligned} K\chi_y(x) &= \sum_z K(x,z)\chi_y(z) \\ &= \left(\frac{1}{n}\sum_i (-1)^{e_i \cdot y}\right)\chi_y(x) = \frac{n - 2|y|}{n}\,\chi_y(x). \end{aligned}$$

This shows that χ_y is an eigenfunction of $I - K$ with eigenvalue $2|y|/n$ where $|y|$ is the number of 1's in y. Thus the eigenvalue $2j/n$ has multiplicity $\binom{n}{j}$ $0 \leq j \leq n$. This information leads to the bound

$$\begin{aligned} \|h_t^x - 1\|_2^2 &= \sum_1^n \binom{n}{j} e^{-4tj/n} \\ &\leq \sum_1^n \frac{n^j}{j!} e^{-4tj/n} \\ &\leq e^{ne^{-4t/n}} - 1. \end{aligned}$$

Hence

$$\|h_t^x - 1\|_2^2 \leq e^{1-c} \quad \text{for} \quad t = \frac{1}{4}\,n\,(\log n\ + c)\,, \quad c > 0.$$

It follows that $T_2(K, 1/e) \leq \frac{1}{4}n(2 + \log n)$. Also, $\|h_t^x - 1\|_2^2 \geq ne^{-4t/n}$ hence $T_2 = T_2(K, 1/e) \geq \frac{1}{4}n(1 + \log n)$. In this case, the lower bound

$$T_2 \geq \frac{1}{\lambda} = \frac{1}{\omega} = \frac{2}{n}$$

is off by a factor of $\log n$ whereas the upper bound

$$T_2 \leq \frac{1}{2\lambda}\left(2 + \log\frac{1}{\pi_*}\right) = \frac{n}{(}2 + n).$$

is off by a factor of $n/\log n$.

2.1.3 Chernoff bounds and central limit theorems

It is well established that ergodic Markov chains satisfy large deviation bounds of Chernoff's type for

$$\mathbf{P}_q\left(\frac{1}{t}\int_0^t f(X_s)ds - \pi(f) > \gamma\right)$$

as well as central limit theorems to the effect that

$$\left|\mathbf{P}_q\left(\int_0^t f(X_s)ds - t\pi(f) \le \sigma t^{1/2}\gamma\right) - \Phi(\gamma)\right| \to 0$$

where $\Phi(\gamma)$ is the cumulative Gaussian distribution and σ is an appropriate number depending on f and K (the asymptotic variance).

The classical treatment of these problems leads to results having a strong asymptotic flavor. Turning these results into quantitative bounds is rather frustrating even in the context of finite Markov chains.

Some progress has been made recently in this direction. This short section presents without any detail two of the main results obtained by Pascal Lezaud [59] and Brad Mann [61] in their Ph.D. theses respectively at Toulouse and Harvard universities.

The work of Lezaud clarifies previous results of Gillman [44] and Dinwoodie [36, 37] on quantitative Chernoff bounds for finite Markov chains. A typical result is as follows (there are also discrete time versions).

Theorem 2.1.8 *Let (K,π) be a finite irreducible Markov chain. Let q denote the initial distribution and $\mathbf{P}_q$ be the law of the associated continuous time process $(X_t)_{t>0}$. Then, for all functions f such that $\pi(f) = 0$ and $\|f\|_\infty \le 1$,*

$$\mathbf{P}_q\left(\frac{1}{t}\int_0^t f(X_s)ds > \gamma\right) \le \|q/\pi\|_2 \exp\left(-\frac{\gamma^2\lambda t}{10}\right).$$

Concerning the Berrry-Essen central limit theorem, we quote a continuous time version of one of Brad Mann's result which has been obtained by Pascal Lezeaud.

Theorem 2.1.9 *Let (K,π) be a finite irreducible reversible Markov chain. Let q denote the initial distribution and $\mathbf{P}_q$ be the law of the associated continuous time process $(X_t)_{t>0}$. Then, for $t > 0$, $-\infty < \gamma < \infty$ and for all functions f such that $\pi(f) = 0$ and $\|f\|_\infty \le 1$,*

$$\left|\mathbf{P}_q\left(\frac{1}{\sigma\sqrt{t}}\int_0^t f(X_s)ds \le \gamma\right) - \Phi(\gamma)\right| \le \frac{100\|q/\pi\|_2\|f\|_2^2}{\lambda^2\,\sigma^3\,t^{1/2}}$$

where

$$\sigma^2 = \lim_{t\to\infty}\frac{1}{t}\mathrm{Var}_\pi\left(\int_0^t f(X_s)ds\right).$$

See [41, 61, 59, 28] for details and examples. There are non-reversible and/or discrete time versions of the last theorem. Mann's Thesis contains a nice discussion of the history of the subject and many references.

2.2 Hypercontractivity

This section introduces the notions of logarithmic Sobolev constant and of hypercontractivity and shows how they enter convergence bounds. A very informative account of the development of hypercontractivity and logarithmic Sobolev inequalities can be found in L. Gross survey paper [47]. See also [7, 8, 15, 16, 46]. The paper [29] develops applications of these notions to finite Markov chains.

2.2.1 The log-Sobolev constant

The definition of the logarithmic Sobolev constant α is similar to that of the spectral gap λ where the variance has been replaced by

$$\mathcal{L}(f) = \sum_{x\in X} |f(x)|^2 \log\left(\frac{|f(x)|^2}{\|f\|_2^2}\right)\pi(x).$$

Observe that $\mathcal{L}(f)$ is nonnegative. This follows from Jensen's inequality applied to the convex function $\phi(t) = t^2\log t^2$. Furthermore $\mathcal{L}(f)=0$ if and only if f is constant.

Definition 2.2.1 *Let K be an irreducible Markov chain with stationary measure π. The logarithmic constant $\alpha=\alpha(K)$ is defined by*

$$\alpha = \min\left\{\frac{\mathcal{E}(f,f)}{\mathcal{L}(f)};\mathcal{L}(f)\neq 0\right\}.$$

It follows from the definition that α is the largest constant c such that the logarithmic Sobolev inequality

$$c\mathcal{L}(f)\le \mathcal{E}(f,f)$$

holds for all functions f. Observe that one can restrict f to be real nonnegative in the definition of α since $\mathcal{L}(f)=\mathcal{L}(|f|)$ and $\mathcal{E}(|f|,|f|)\le\mathcal{E}(f,f)$.

To get a feel for this notion we prove the following result.

Lemma 2.2.2 *For any chain K the log-Sobolev constant α and the spectral gap λ satisfy $2\alpha\le\lambda$.*

PROOF: We follow [67]. Let g be real and set $f=1+\varepsilon g$ and write, for ε small enough

$$\begin{aligned} |f|^2\log|f|^2 &= 2\left(1+2\varepsilon g+\varepsilon^2|g|^2\right)\left(\varepsilon g-\frac{\varepsilon^2|g|^2}{2}+O(\varepsilon^3)\right)\\ &= 2\varepsilon g+3\varepsilon^2|g|^2+O(\varepsilon^3)\end{aligned}$$

and

$$\begin{aligned} |f|^2\log\|f\|_2^2 &= \left(1+2\varepsilon g+\varepsilon^2|g|^2\right)\left(2\varepsilon\pi(g)+\varepsilon^2\|g\|_2^2-2\varepsilon^2(\pi(g))^2+O(\varepsilon^3)\right)\\ &= 2\varepsilon\pi(g)+4\varepsilon^2 g\pi(g)+\varepsilon^2\|g\|_2^2-2\varepsilon^2(\pi(g))^2+O(\varepsilon^3).\end{aligned}$$

Thus,

$$|f|^2 \log \frac{|f|^2}{\|f\|_2^2} = 2\varepsilon(g - \pi(g)) + \varepsilon^2 \left(3|g|^2 - \|g\|_2^2 - 4g\pi(g) + 2(\pi(g))^2\right) + O(\epsilon^3)$$

and

$$\begin{aligned}\mathcal{L}(f) &= 2\varepsilon^2 \left(\|g\|^2 - (\pi(g))^2\right) + O(\varepsilon^3) \\ &= 2\varepsilon^2 \mathrm{Var}(g) + O(\varepsilon^3).\end{aligned}$$

To finish the proof, observe that $\mathcal{E}(f,f) = \varepsilon^2 \mathcal{E}(g,g)$, multiply by ε^{-2}, use the variational characterizations of α and λ, and let ε tend to zero.

It is not completely obvious from the definition that $\alpha(K) > 0$ for any finite irreducible Markov chain. The next result, adapted from [65, 66, 67], yields a proof of this fact.

Theorem 2.2.3 *Let K be an irreducible Markov chain with stationary measure π. Let α be its logarithmic Sobolev constant and λ its spectral gap. Then either $\alpha = \lambda/2$ or there exists a positive non-constant function u which is solution of*

$$2u \log u - 2u \log \|u\|_2 - \frac{1}{\alpha}(I - K)u = 0, \tag{2.2.1}$$

and such that $\alpha = \mathcal{E}(u,u)/\mathcal{L}(u)$. In particular $\alpha > 0$.

Proof: Looking for a minimizer of $\mathcal{E}(f,f)/\mathcal{L}(f)$, we can restrict ourselves to non-negative functions satisfying $\pi(f) = 1$. Now, either there exists a non-constant non-negative minimizer (call it u), or the minimum is attained at the constant function 1 where $\mathcal{E}(1,1) = \mathcal{L}(1) = 0$. In this second case, the proof of Lemma 2.2.2 shows that we must have $\alpha = \lambda/2$ since, for any function $g \not\equiv 0$ satisfying $\pi(g) = 0$,

$$\lim_{\varepsilon \to 0} \frac{\mathcal{E}(1+\varepsilon g, 1+\varepsilon g)}{\mathcal{L}(1+\varepsilon g)} = \lim_{\varepsilon \to 0} \frac{\varepsilon^2 \mathcal{E}(g,g)}{2\varepsilon^2 \mathrm{Var}_\pi(g)} \geq \frac{\lambda}{2}.$$

Hence, either $\alpha = \lambda/2$ or there must exist a non-constant non-negative function u which minimizes $\mathcal{E}(f,f)/\mathcal{L}(f)$. It is not hard to show that any minimizer of $\mathcal{E}(f,f)/\mathcal{L}(f)$ must satisfy (2.2.1). Finally, if $u \geq 0$ is not constant and satisfies (2.2.1) then u must be positive. Indeed, if it vanishes at $x \in \mathcal{X}$ then $Ku(x) = 0$ and u must vanishe at all points y such that $K(x,y) > 0$. By irreducibility, this would imply $u \equiv 0$, a contradiction.

2.2.2 Hypercontractivity, α, and ergodicity

We now recall the main result relating log-Sobolev inequalities to the so-called hypercontractivity of the semigroup H_t. For a history of this result see Gross' survey [47]. See also [7, 8, 16, 46]. A proof can also be found in [29].

Theorem 2.2.4 *Let (K,π) be a finite Markov chain with log-Sobolev constant α.*

1. *Assume that there exists $\beta > 0$ such that $\|H_t\|_{2\to q} \le 1$ for all $t > 0$ and $2 \le q < +\infty$ satisfying $e^{4\beta t} \ge q-1$. Then $\beta\mathcal{L}(f) \le \mathcal{E}(f,f)$ for all f and thus $\alpha \ge \beta$.*
2. *Assume that (K,π) is reversible. Then $\|H_t\|_{2\to q} \le 1$ for all $t > 0$ and all $2 \le q < +\infty$ satisfying $e^{4\alpha t} \ge q-1$.*
3. *For non-reversible chains, we still have $\|H_t\|_{2\to q} \le 1$ for all $t > 0$ and all $2 \le q < +\infty$ satisfying $e^{2\alpha t} \ge q-1$.*

We will not prove this result but only comment on the different statements. First let us assume that (K,π) is reversible. The first two statements show that α can also be characterized as the largest β such that

$$\|H_t\|_{2\to q} \le 1 \text{ for all } t > 0 \text{ and all } 2 \le q < +\infty \text{ satisfying } e^{4\beta t} \ge q-1. \tag{2.2.2}$$

Recall that H_t is always a contraction on $\ell^2(\pi)$ and that, in fact, $\|H_t\|_{2\to 2} = 1$ for all $t > 0$. Also, (1.3.2) and (1.3.5) easily show that $\|H_t\|_{2\to\infty} > 1$ for all $t > 0$ and tends to 1 as t tends to infinity. Thus, even in the finite setting, it is rather surprising that for each $2 < q < \infty$ there exists a finite $t_q > 0$ such that $\|H_t\|_{2\to q} \le 1$ for $t \ge t_q$. The fact that such a t_q exists follows from Theorem 2.2.3 and Theorem 2.2.4(2).

Statements 2 and 3 in Theorem 2.2.4 are the keys of the following theorem which describes how α enters quantitative bounds on convergence to stationarity.

Theorem 2.2.5 *Let (K,π) be a finite Markov chain. Then, for $\varepsilon,\theta,\sigma \ge 0$ and $t = \varepsilon+\theta+\sigma$,*

$$\|h_t^x - 1\|_2 \le \begin{cases} \|h_\varepsilon^x\|_2^{2/(1+e^{4\alpha\theta})}\, e^{-\lambda\sigma} & \textit{if } (K,\pi) \textit{ is revesible} \\ \|h_\varepsilon^x\|_2^{2/(1+e^{2\alpha\theta})}\, e^{-\lambda\sigma} & \textit{in general.} \end{cases} \tag{2.2.3}$$

In particular,

$$\|h_t^x - 1\|_2 \le e^{1-c} \tag{2.2.4}$$

for all $c \ge 0$ and

$$t = \begin{cases} (4\alpha)^{-1}\log_+\log(1/\pi(x)) + \lambda^{-1}c & \textit{for reversible chains} \\ (2\alpha)^{-1}\log_+\log(1/\pi(x)) + \lambda^{-1}c & \textit{in general} \end{cases}$$

where $\log_+ t = \max\{0, \log t\}$.

PROOF: We treat the general case. The improvement for reversible chains follows from Theorem 2.2.4(2). For $\theta > 0$, set $q(\theta) = 1 + e^{2\alpha\theta}$. The third statement of Theorem 2.2.4(3) gives $\|H_\theta\|_{2\to q(\theta)} \le 1$. By duality, it follows that $\|H_\theta^*\|_{q'(\theta)\to 2} \le 1$ where $q'(\theta)$ is the Hölder conjugate of $q(\theta)$ defined by $1/q'(\theta)+1/q(\theta)=1$. Write

$$\begin{aligned}\|h^x_{\varepsilon+\theta+\sigma}-1\|_2 &= \|(H^*_{\theta+\sigma}-\pi)h^x_\varepsilon\|_2 \le \|H^*_\theta h^x_\varepsilon\|_2\|H^*_\sigma-\pi\|_{2\to 2}\\ &\le \|h^x_\varepsilon\|_{q'(\theta)}\|H^*_\theta\|_{q'(\theta)\to 2}\|H^*_\sigma-\pi\|_{2\to 2} = \|h^x_\varepsilon\|_2^{2/q(\theta)}\, e^{-\lambda\sigma}.\end{aligned}$$

Here we have used $1\le q'\le 2$ and the Hölder inequality

$$\|f\|_{q'} \le \|f\|_1^{1-2/q}\|f\|_2^{2/q}$$

with $f=h^x_\varepsilon$, $\|h^x_\varepsilon\|_1=1$ to obtain the last inequality.

Consider the function δ_x defined by $\delta_x(x)=1/\pi(x)$ and $\delta_x(y)=0$ for $x\neq y$ and observe that $h^x_0=\delta_x$, $\|h^x_0\|_2=\|\delta_x\|_2\le 1/\pi(x)^{1/2}$. Hence, for $t=\theta+\sigma$,

$$\|h^x_t-1\|_2 \le \left(\frac{1}{\pi(x)}\right)^{1/(1+e^{2\alpha\theta})} e^{-\lambda\sigma}.$$

Assuming $\pi(x)<1/e$ and choosing

$$\theta=\frac{1}{2\alpha}\log\log\frac{1}{\pi(x)},\quad \sigma=\frac{c}{\lambda}$$

we obtain $\|h_t-1\|_2\le e^{1-c}$ which is the desired inequality. When $\pi(x)\ge 1/e$, simply use $\theta=0$.

Corollary 2.2.6 *Let (K,π) be a finite Markov chain. Then*

$$\left|\frac{H_t(x,y)}{\pi(y)}-1\right| = |h_t(x,y)-1| \le e^{2-c} \tag{2.2.5}$$

for all $c>0$ and

$$t=\begin{cases}(4\alpha)^{-1}\big(\log_+\log(1/\pi(x))+\log_+\log(1/\pi(y))\big)+\lambda^{-1}c & \textit{(reversible)}\\ (2\alpha)^{-1}\big(\log_+\log(1/\pi(x))+\log_+\log(1/\pi(y))\big)+\lambda^{-1}c & \textit{(general).}\end{cases}$$

PROOF: Use Theorem 2.2.5 for both H_t and H_t^* together with

$$|h_{t+s}(x,y)-1| \le \|h^x_t-1\|_2\|h^{*y}_s-1\|_2.$$

The next result must be compared with Theorem 2.1.7.

Corollary 2.2.7 *Let (K,π) be a finite reversible Markov chain. For $1\le p\le\infty$, let T_p be defined by* (2.1.3)*. Then, for $1\le p\le 2$,*

$$\frac{1}{2\alpha}\le T_p\le\frac{1}{4\alpha}\left(4+\log_+\log\frac{1}{\pi_*}\right)$$

and for $2<p\le\infty$,

$$\frac{1}{2\alpha}\le T_p\le\frac{1}{2\alpha}\left(3+\log_+\log\frac{1}{\pi_*}\right)$$

where $\pi_*=\min_x\pi(x)$ as in (1.4.2)*. Similar upper bounds holds in the non-reversible case (simply multiply the right-hand side by* 2*).*

This result shows that α is closely related to the quantity we want to bound, namely the "time to equilbrium" T_2 (more generally T_p) of the chain (K, π). The natural question now is:

$$\boxed{\text{can one compute or estimate the constant } \alpha?}$$

Unfortunately, the present answer is that it seems to be a very difficult problem to estimate α. To illustrate this point we now present what, in some sense, is the only example of finite Markov chain for which α is known explicitely.

EXAMPLE 2.2.1: Let $\mathcal{X} = \{0,1\}$ be the two point space. Fix $0 < \theta \leq 1/2$. Consider the Markov kernel $K = K_\theta$ given by $K(0,0) = K(1,0) = \theta$, $K(0,1) = K(1,1) = 1-\theta$. The chain K_θ is reversible with respect to π_θ where $\pi_\theta(0) = (1-\theta)$, $\pi_\theta(1) = \theta$.

Theorem 2.2.8 *The log-Sobolev constant of the chain (K_θ, π_θ) on $\mathcal{X} = \{0,1\}$ is given by*

$$\alpha_\theta = \frac{1-2\theta}{\log[(1-\theta)/\theta]}$$

with $\alpha_{1/2} = 1/2$.

PROOF: The case $\theta = 1/2$ is due to Aline Bonami [10] and is well known since the work of L. Gross [46]. The case $\theta < 1/2$ has only been worked out recently in [29] and independently in [48]. The present elegant proof is due to Sergei Bobkov. He kindly authorized me to include his argument in these notes.

First, linearize the problem by observing that

$$\mathcal{L}(f) = \sup\left\{\langle f^2, g\rangle : g \neq 0,\ \|e^g\|_1 = 1\right\}.$$

Hence

$$\alpha = \inf\{\alpha(g) : g \neq 0,\ \|e^g\|_1 = 1\}$$

with

$$\alpha(g) = \inf\left\{\frac{\mathcal{E}_\theta(f,f)}{\langle f^2, g\rangle} : f \neq 0\right\}$$

where $\mathcal{E}_\theta$ is the Dirichlet form $\mathcal{E}_\theta(f,f) = \theta(1-\theta)|f(0) - f(1)|^2$. This is valid for any Markov chain.

We now return to the two point space. Fix $g \neq 0$ and set $g(0) = b$, $g(1) = a$ with $\theta e^a + (1-\theta)e^b = 1$. Observe that this implies $ab < 0$. To find $\alpha_\theta(g)$ we can assume $f > 0$, $f(0) = \sqrt{x}$, $f(1) = \sqrt{y} = 1$ with $x > 0$. Then

$$\alpha_\theta(g) = \inf_{x>0}\left\{\frac{\theta(1-\theta)(\sqrt{x}\ -1)^2}{\theta x a + (1-\theta)b}\right\}.$$

One easily checks that the infimum is attained for $x = [(1-\theta)b/\theta a]^2$. Therefore

$$\alpha_\theta(g) = \frac{\theta}{b} + \frac{1-\theta}{a}.$$

It follows that

$$\alpha_\theta = \inf\left\{\frac{\theta}{b} + \frac{1-\theta}{a} : \theta e^a + (1-\theta)e^b = 1\right\}.$$

We set

$$t = e^a, \quad s = e^b$$

and

$$h(t) = \frac{\theta}{\log s} + \frac{1-\theta}{\log t} \quad \text{with} \quad \theta t + (1-\theta)s = 1,$$

so that

$$\alpha_\theta = \inf\left\{h(t) : t \in (0,1) \cup (1, 1/\theta)\right\}.$$

By Taylor expansion at $t = 1$,

$$h(t) = \frac{1}{2} + \frac{2\theta - 1}{12(1-\theta)}(t-1) + \frac{\theta^3 + (1-\theta)^3}{24(1-\theta)^2}(t-1)^2 + O((t-1)^3).$$

So, we extend h as a continuous function on $[0, 1/\theta]$ by setting

$$h(0) = -\theta/\log(1-\theta), \;\; h(1) = 1/2, \;\; h(1/\theta) = -(1-\theta)/\log\theta.$$

Observe that $h(1)$ is not a local minimum if $\theta \neq 1/2$. We have

$$h'(t) = \frac{\theta^2}{(1-\theta)s[\log s]^2} - \frac{(1-\theta)}{t[\log t]^2}.$$

This shows that neither $h(0)$ nor $h(1/\theta)$ are minima of h since $h'(0) = -\infty$, $h'(1/\theta) = +\infty$.

Let us solve $h'(t) = 0$ and show that this equation has a unique solution in $(0, 1/\theta)$. The condition $h'(t) = 0$ is equivalent to (recall that $(\log s)(\log t) < 0$)

$$\begin{cases} \theta\sqrt{t}\log t = -(1-\theta)\sqrt{s}\log s \\ \theta t + (1-\theta)s = 1 \end{cases}.$$

Since $\theta t + (1-\theta)s = 1$, we have $\theta = (1-s)/(t-s)$, $1 - \theta = (1-t)/(s-t)$. Hence $h'(t) = 0$ implies $s = t = 1$ or

$$\frac{\sqrt{t}\log t}{1-t} = \frac{\sqrt{s}\log s}{1-s}.$$

The function $t \to v(t) = \frac{\sqrt{t}\log t}{1-t}$ satisfies $v(0) = v(+\infty) = 0$, $v(1) = -1$ and $v(1/t) = v(t)$. It is decreasing on $(0,1)$ and increasing on $(1, +\infty)$. It follows that $h'(t) = 0$ implies that either $s = t = 1$ or $t = 1/s = (1-\theta)/\theta$ (because $\theta t + (1-\theta)s = 1$). If $\theta \neq 1/2$ then $h'(1) \neq 0$, the equation $h'(t) = 0$ has a unique solution $t = (1-\theta)/\theta$ and

$$\min_{t\in(0,1/\theta)} h(t) = h((1-\theta)/\theta) = \frac{1-2\theta}{\log[(1-\theta)/\theta]}.$$

If $\theta = 1/2$, then $h'(1) = 0$ and 1 is the only solution of $h'(t) = 0$ so that $\min_{t\in(0,2)} h(t) = h(1) = 1/2$ in this case. This proves Theorem 2.2.8.

EXAMPLE 2.2.2: Using Theorems 2.2.3 and 2.2.8, one obtains the following result.

Theorem 2.2.9 *Let π be a positive probability measure on $\mathcal{X}$. Let $K(x,y) = \pi(y)$. Then the log-Sobolev constant of (K,π) is given by*

$$\alpha = \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]}$$

where $\pi_* = \min_{\mathcal{X}} \pi$.

PROOF: Theorem 2.2.3 shows that any non trivial minimizer must take only two values. The desired result then follows from Theorem 2.2.8. See [29] for details. THeorem 2.2.9 yields a sharp universal lower bound on α in terms of λ.

Corollary 2.2.10 *The log-Sobolev constant α and the spectral gap λ of any finite Markov chain K with stationary measure π satisfy*

$$\alpha \geq \frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]}\,\lambda.$$

PROOF: The variance $\mathrm{Var}_\pi(f)$ is nothing else than the Dirichlet form of the chain considered in Theorem 2.2.9. Hence

$$\frac{1-2\pi_*}{\log[(1-\pi_*)/\pi_*]}\,\mathcal{L}_\pi(f) \leq \mathrm{Var}_\pi(f) \leq \frac{1}{\lambda}\mathcal{E}_{K,\pi}(f,f).$$

The desired result follows.

2.2.3 Some tools for bounding α from below

The following two results are extremely useful in providing examples of chains where α can be either computed or bounded from below. Lemma 2.2.11 computes the log-Sobolev constant of products chains. This important result is due (in greater generality) to I Segal and to W. Faris, see [47]. Lemma 2.2.12 is a comparison result.

Lemma 2.2.11 *Let (K_i,π_i), $i = 1,\dots,d$, be Markov chains on finite sets $\mathcal{X}_i$ with spectral gaps λ_i and log-Sobolev constants α_i. Fix $\mu = (\mu_i)_1^d$ such that $\mu_i > 0$ and $\sum \mu_i = 1$. Then the product chain (K,π) on $\mathcal{X} = \prod_1^d \mathcal{X}_i$ with Kernel*

$$\begin{aligned} K_\mu(x,y) &= K(x,y) \\ &= \sum_1^d \mu_i \delta(x_1,y_1)\dots\delta(x_{i-1},y_{y-1})K_i(x_i,y_i)\delta(x_{i+1},y_{i+1})\dots\delta(x_d,y_d) \end{aligned}$$

(where $\delta(x,y)$ vanishes for $x \neq y$ and $\delta(x,x) = 1$) and stationary measure $\pi = \bigotimes_1^d \pi_i$ satisfies

$$\lambda = \min_i \{\mu_i \lambda_i\} \ , \quad \alpha = \min_i \{\mu_i \alpha_i\}.$$

PROOF: Let $\mathcal{E}_i$ denote the Dirichlet form associated to K_i, then the product chain K has Dirichlet form

$$\mathcal{E}(f,f) = \sum_1^d \mu_i \left(\sum_{x_j : j \neq i} \mathcal{E}_i(f,f)(x^i) \pi^i(x^i) \right)$$

where x^i is the sequence $(x_1, \dots, x_d)$ with x_i omitted, $\pi^i = \bigotimes_{\ell : \ell \neq i} \pi_\ell$ and $\mathcal{E}_i(f,f)(x^i) = \mathcal{E}_i(f(x_1, \dots, x_d), f(x_1, \dots, x_d))$ has the obvious meaning: $\mathcal{E}_i$ acts on the i^{th} coordinate whereas the other coordinates are fixed. It is enough to prove the Theorem when $d = 2$. We only prove the statement for α. The proof for λ is similar. Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ be a nonnegative function and set $F(x_2) = \left(\sum_{x_1} f(x_1, x_2)^2 \pi_1(x_1) \right)^{1/2}$. Write

$$\begin{aligned}
\mathcal{L}(f) &= \sum_{x_1, x_2} |f(x_1,x_2)|^2 \log \frac{f(x_1,x_2)^2}{\|f\|_{2,\pi}^2} \, \pi(x_1,x_2) \\
&= \sum_{x_2} |F(x_2)|^2 \log \frac{F(x_2)^2}{\|F\|_{2,\pi_2}^2} \, \pi_2(x_2) \\
&\quad + \sum_{x_1,x_2} |f(x_1,x_2)|^2 \log \frac{f(x_1,x_2)^2}{F(x_2)^2} \, \pi(x_1,x_2) \\
&\leq [\mu_2 \alpha_2]^{-1} \mu_2 \mathcal{E}_2(F,F) + [\mu_1 \alpha_1]^{-1} \sum_{x_2} \mu_1 \mathcal{E}_1(f(\cdot, x_2), f(\cdot, x_2)) \, \pi_2(x_2).
\end{aligned}$$

Now, the triangle inequality

$$\begin{aligned}
|F(x_2) - F(y_2)| &= | \, \|f(\cdot, x_2)\|_{2,\pi_1} - \|f(\cdot, y_2)\|_{2,\pi_1} \, | \\
&\leq \|f(\cdot, x_2) - f(\cdot, y_2)\|_{2,\pi_1}
\end{aligned}$$

implies that

$$\mathcal{E}_2(F,F) \leq \sum_{x_1} \mathcal{E}_2(f(x_1, \cdot), f(x_1, \cdot)) \, \pi_1(x_1).$$

Hence

$$\begin{aligned}
\mathcal{L}(f) \leq\ & [\mu_2 \alpha_2]^{-1} \sum_{x_1} \mu_2 \mathcal{E}_2(f(x_1, \cdot), f(x_1, \cdot)) \, \pi_1(x_1) \\
& + [\mu_1 \alpha_1]^{-1} \sum_{x_2} \mu_1 \mathcal{E}_1(f(\cdot, x_2), f(\cdot, x_2)) \pi_2(x_2)
\end{aligned}$$

which yields

$$\mathcal{L}(f) \leq \max_i \{1/[\mu_i \alpha_i]\} \mathcal{E}(f,f).$$

This shows that $\alpha \geq \min_i[\mu_i \alpha_i]$. Testing on functions that depend only on one of the two variables shows that $\alpha = \min_i[\mu_i \alpha_i]$.

EXAMPLE 2.2.3: Fix $0 < \theta < 1$. Take each $\mathcal{X}_i = \{0,1\}$, $\mu_i = 1/d$, $K_i = K_\theta$ as in Theorem 2.2.8. We obtain a chain on $\mathcal{X} = \{0,1\}^d$ which proceeds as follows. If the current state is x, we pick a coordinate, say i, uniformly at random. If $x_i = 0$ we change it to 1 with probability $1-\theta$ and do nothing with probability θ. If $x_i = 1$ we change it to 0 with probability θ and do nothing with pobability $1-\theta$. According to Lemma 2.2.11, this chain has spectral gap $\lambda = 1/d$ and log-Sobolev constant

$$\alpha = \frac{1-2\theta}{d \log[(1-\theta)/\theta]}.$$

Observe that the function $F(t) : t \to c(1-\theta-t)$ with $c = (\theta(1-\theta))^{-1/2}$ is an eigenfunction of K_i (for each i) with eigenvalue $0 = 1-\lambda$ satisfying $\|F_i\|_2 = 1$. It follows that the eigenvalues of $I-K$ are the numbers j/d each with multiplicity $\binom{d}{j}$. The corresponding orthonormal eigenfunctions are

$$F_I : (x)_1^d \to \prod_{i \in I} F_i(x)$$

where $I \subset \{1, \ldots, d\}$, $F_i(x) = F(x_i)$ and $\#I = j$. The product structure of the chain K yields

$$\|h_t^x - 1\|_2^2 = h_{2t}(x,x) - 1 = \prod_1^d (1 + |F_i(x)|^2 e^{-2t/d})^d - 1.$$

For instance,

$$\begin{aligned} \|h_t^{\mathbf{0}} - 1\|_2^2 &= \left(1 + \frac{1-\theta}{\theta}\, e^{-2t/d}\right)^d - 1 \\ &\leq \frac{(1-\theta)d}{\theta}\, e^{-2t/d} e^{\frac{(1-\theta)d}{\theta}\, e^{-2t/d}}. \end{aligned}$$

In particular

$$\|h_t^{\mathbf{0}} - 1\|_2 \leq e^{\frac{1}{2}-c} \quad \text{for} \quad t = \frac{d}{2}\left(\log[(1-\theta)d/\theta] + 2c\right), \quad c > 0.$$

Hence

$$T_2(K_\theta, 1/e) \leq \frac{d}{2}\left(3 + \log[(1-\theta)d/\theta]\right), \quad c > 0.$$

Also, we have

$$\|h_t^{\mathbf{0}} - 1\|_2^2 \geq \frac{(1-\theta)d}{\theta} e^{-2t/d}$$

which shows that the upper bound obtained above is sharp and that

$$T_2(K, 1/e) \geq \frac{d}{2}\left(2 + \log[(1-\theta)d/\theta]\right).$$

It is instructive to compare these precise results with the upper bound which follows from Theorem 2.2.5. In the present case this theorem yields

$$\|h_t^{\mathbf{0}} - 1\|_2 \leq e^{1-c} \quad \text{for } t = \frac{d}{2}\left(\frac{1}{2(1-2\theta)}\left(\log\frac{1-\theta}{\theta}\right)\log d \; + 2c\right).$$

For any fixed $\theta < 1/2$, this is slightly off, but of the right order of magnitude. For $\theta = 1/2$ this simplifies to

$$\|h_t^{\mathbf{0}} - 1\|_2 \leq e^{1-c} \quad \text{for } t = \frac{d}{2}(\log d \; + 2c)$$

which is very close to the sharp result described above. In this case, the upper bound

$$T_2 = T_2(K_{1/2}, 1/e) \leq \frac{1}{4\alpha}\left(4 + \log_+ \log\frac{1}{\pi_*}\right) \leq \frac{d}{2}\left(4 + \log d\right)$$

of Corollary 2.2.7 compares well with the lower bound

$$T_2 \geq \frac{d}{2}(2 + \log d).$$

Example 2.2.4: Consider now $|x| = \sum_1^d x_i$, that is, the number of 1's in the chain in the preceding example, as random variable taking values in $\mathcal{X}_0 = \{0, \ldots, d\}$. Clearly, this defines a Markov chain on $\mathcal{X}_0$ with stationary measure

$$\pi_0(j) = \theta^j (1-\theta)^{d-j}\binom{d}{j}$$

and kernel

$$K_0(i,j) = \begin{cases} 0 & \text{if } |i-j| > 1 \\ (1-\theta)(1-i/d) & \text{if } j = i+1 \\ \theta i/d & \text{if } j = i-1 \\ (1-\theta)i/d + \theta(1-i/d) & \text{if } i = j. \end{cases}$$

All the eigenvalues of $I - K_0$ are also eigenvalues of $I - K$. It follows that $\lambda_0 \geq 1/d$. Furthermore, the function $F : i \to c_0[d(1-\theta) - i]$ with $c_0 = (d\theta(1-\theta))^{-1/2}$ is an eigenfunction with eigenvalue $1/d$ and $\|F\|_2 = 1$. Hence, $\lambda_0 = 1/d$. Concerning α_0, all we can say is that

$$\alpha_0 \geq \frac{1-2\theta}{d\log[(1-\theta)/\theta]}.$$

When $\theta = 1/2$ this inequality and Lemma 2.2.2 show that $\alpha_0 = 1/(2d) = \lambda/2$.

The next result allows comparison of the spectral gaps and log-Sobolev constants of two chains defined on different state spaces.

Lemma 2.2.12 *Let (K,π), (K',π') be two Markov chains defined respectively on the finite sets $\mathcal{X}$ and $\mathcal{X}'$. Assume that there exists a linear map*

$$\ell^2(\mathcal{X},\pi) \to \ell^2(\mathcal{X}',\pi') : f \to \tilde{f}$$

and constants $A, B, a > 0$ such that, for all $f \in \ell^2(\mathcal{X},\pi)$

$$\mathcal{E}'(\tilde{f},\tilde{f}) \le A\mathcal{E}(f,f) \quad \text{and} \quad a\mathrm{Var}_\pi(f) \le \mathrm{Var}_{\pi'}(\tilde{f}) + B\mathcal{E}(f,f)$$

then

$$\frac{a\lambda'}{A + B\lambda'} \le \lambda\ .$$

Similarly, if

$$\mathcal{E}'(\tilde{f},\tilde{f}) \le A\mathcal{E}(f,f) \quad \text{and} \quad a\mathcal{L}_\pi(f) \le \mathcal{L}_{\pi'}(\tilde{f}) + B\mathcal{E}(f,f),$$

then

$$\frac{a\alpha'}{A + B\alpha'} \le \alpha.$$

In particular, if $\mathcal{X} = \mathcal{X}'$, $\mathcal{E}' \le A\mathcal{E}$ and $a\pi \le \pi'$, then

$$\frac{a\lambda'}{A} \le \lambda, \qquad \frac{a\alpha'}{A} \le \alpha.$$

Proof: The two first assertions follow from the variational definitions of λ and α. For instance, for λ we have

$$\begin{aligned} a\mathrm{Var}_\pi(f) &\le \mathrm{Var}_{\pi'}(\tilde{f}) + B\mathcal{E}(f,f) \\ &\le \frac{1}{\lambda'}\mathcal{E}'(\tilde{f},\tilde{f}) + B\mathcal{E}(f,f) \\ &\le \left(\frac{A}{\lambda'} + B\right)\mathcal{E}(f,f). \end{aligned}$$

The desired inequality follows.

To prove the last assertion, use $a\pi \le \pi'$ and the formula

$$\mathrm{Var}_\pi(f) = \min_{c\in\mathbb{R}} \sum_x |f(x) - c|^2\pi(x)$$

to see that $a\mathrm{Var}_\pi(f) \le \mathrm{Var}_{\pi'}(f)$. The inequality between log-Sobolev constants follows from $\xi\log\xi - \xi\log\zeta - \xi + \zeta \ge 0$ for all $\xi, \zeta > 0$ and

$$\begin{aligned} \mathcal{L}_\pi(f) &= \sum_x \left(|f(x)|^2\log|f(x)|^2 - |f(x)|^2\log\|f\|_2^2 - |f(x)|^2 + \|f\|_2^2\right)\pi(x) \\ &= \min_{c>0}\sum_x \left(|f(x)|^2\log|f(x)|^2 - |f(x)|^2\log c - |f(x)|^2 + c\right)\pi(x). \end{aligned}$$

This useful observation is due to Holley and Stroock [50].

EXAMPLE 2.2.5: Let $\mathcal{X} = \{0,1\}^n$ and set $|x-y| = \sum_i |x_i - y_i|$. Let $\tau : \mathcal{X} \to \mathcal{X}$ be the map defined by $\tau(x) = y$ where $y_i = x_{i-1}$, $1 < i \le n$, $y_1 = x_n$. Consider the chain

$$K(x,y) = \begin{cases} 1/(n+1) & \text{if } |x-y| = 1 \\ 1/(n+1) & \text{if } y = \tau(x) \\ 0 & \text{oherwise.} \end{cases}$$

It is not hard to check that the uniform distribution $\pi \equiv 2^{-n}$ is the stationary measure of K. Observe that K is neither reversible nor an invariant chain on the group $\{0,1\}^n$. We will study this chain by comparison with the classic chain K' whose kernel vanishes if $|x-y| \neq 1$ and is equal to $1/n$ if $|x-y| = 1$. These two chains have the same stationary measure $\pi \equiv 2^{-n}$. Obviously the Dirichlet forms $\mathcal{E}'$ and $\mathcal{E}$ satisfy

$$\mathcal{E}' \le \frac{n+1}{n} \mathcal{E}(f,f).$$

Applying Lemma 2.2.12, and using the known values $\lambda' = 2/n$, $\alpha' = 1/n$ of the spectral gap and log Sobolev constant of the chain K', we get

$$\lambda \ge \frac{2}{n+1}, \quad \alpha \ge \frac{1}{n+1}.$$

To obtain upper bounds, we use the test function $f = \sum_i (x_i - 1/2)$. This has $\pi(f) = 0$. Also

$$\mathcal{E}(f,f) = \frac{n}{n+1} \mathcal{E}'(f,f) = \frac{n}{n+1} \frac{2}{n} \operatorname{Var}_\pi(f).$$

The first equality follows from the fact that $f(\tau(x)) = f(x)$. The second follows from the fact that f is an eigenvalue of $I - K'$ associated with the eigenvalue $2/n$ (in fact, one can check that f is an eigenfunction of K itself). Hence $\lambda \le 2/(n+1)$. This implies

$$\lambda = \frac{2}{n+1}, \quad \alpha = \frac{1}{n+1}.$$

Applying Theorem 2.2.5 we get

$$\|h_t^x - 1\|_2 \le e^{1-c} \quad \text{for} \quad t = \frac{n+1}{4}\left(2c + \log n\right), \;\; c > 0.$$

The test function f used above has $\|f\|_\infty = n/2$ and $\|f\|_2^2 = n/4$ and is an eigenfunction associated with λ. Hence

$$\max_x \|h_t^x - 1\|_2 = \|H_t - \pi\|_{2\to\infty} \ge \frac{\|H_t f\|_\infty}{\|f\|_2} = n^{1/2} e^{-2t/(n+1)}.$$

This proves the sharpness of our upper bound. A lower bound in ℓ^1 can be obtained by observing that the number of 1's in x, that is $|x|$, evolves has a Markov chain on $\{0,\dots,n\}$ which is essentially the classic Ehrenfest's urn Markov chain.

This example generalizes easily as follows. The permutaion τ can be replaced by any other permutation without affecting the analysis presented above. We can also pick at random among several permutations of the coordinates. This will simply change the factor of comparison between $\mathcal{E}$ and $\mathcal{E}'$.

We end this section with a result that bounds α in terms of $\max_x \|h_t^x - 1\|_2 = \|H_t - \pi\|_{2\to\infty}$. See [29] for a proof. Similar results can be found in [8, 16]

Theorem 2.2.13 *Assume that (K, π) is reversible. Fix $2 < q \le +\infty$ and assume that t_q, M_q satisfy $\|H_{t_q} - \pi\|_{2\to q} \le M_q$. Then*

$$\alpha \ge \frac{(1 - \frac{2}{q})\lambda}{2(\lambda t_q + \log M_q + \frac{q-2}{q})} .$$

In particular, if $q = \infty$ and t is such that $\max_x \|h_t^x - 1\|_2 \le M$, we have

$$\alpha \ge \frac{\lambda}{2(\lambda t + \log M)} .$$

EXAMPLE 2.2.6: Consider the nearest neighbor chain K on $\{0, \dots, n\}$ with loops at the ends. Then $\lambda = 1 - \cos \frac{\pi}{n+1}$. At the end of Section 2.1 it is proved that

$$\|H_t - \pi\|_{2\to\infty}^2 = \max_x \|h_t^x - 1\|_2^2 \le 2e^{-4t/(n+1)^2} \left(1 + \sqrt{(n+1)^2/4t}\right).$$

Thus, for $t = \frac{1}{2}(n+1)^2$, $\|H_t - \pi\|_{2\to\infty} \le 1$. Using this and $\lambda \ge 2/(n+1)^2$ in Theorem 2.2.13 give

$$\frac{1}{2(n+1)^2} \le \alpha \le \frac{1}{2}\left(1 - \cos\frac{\pi}{n+1}\right) = \frac{\pi^2}{4(n+1)^2} + O(1/n^4).$$

The exact value of α is not known.

2.3 Nash inequalities

A Nash inequality for the finite Markov chain (K, π) is an inequality of the type

$$\forall\, f \in \ell^2(\mathcal{X}, \pi), \quad \|f\|_2^{2(1+2/d)} \le C\left(\mathcal{E}(f,f) + \frac{1}{T}\|f\|_2^2\right)\|f\|_1^{4/d}$$

where d, C, T are constants depending on K. The size of these constants is of course crucial in our applications. This inequality implies (in fact, is equivalent to)

$$H_t(x,y) \le B(d)\pi(y)\,(C/t)^{d/2} \ \text{ for } \ 0 < t \le T$$

where $B(d)$ depends only on d and d, C, T are as above. This is discussed in detail in this section. Nash inequalities have received considerable attention in recent years. I personally learned about them from Varopoulos [78]. Their use is emphasized in [11]. Applications to finite Markov chains are presented in [28], with many examples. See also [69]

2.3.1 Nash's argument for finite Markov chains I

Nash introduced his inequality in [64] to study the decay of the heat kernel of certain parabolic equations in Euclidean space. His argument only uses the formula 2.1.2 for the time derivative of $u(t) = \|H_t f\|_2^2$ which reads $u'(t) = -2\mathcal{E}(H_t f, H_t)$. This formula shows that any functional inequality between the ℓ^2 norm of g and the Dirichlet form $\mathcal{E}(g,g)$ (for all g, thus $g = H_t f$) can be translated into a differential inequation involving u. Namely, assume that the Dirichlet form $\mathcal{E}$ satisfies the inequality

$$\forall\, g,\ \ \mathrm{Var}_\pi(g)^{1+2/d} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

Then fix f satisfying $\|f\|_1 = 1$ and set $u(t) = \|H_t(f-\pi(f))\|_2^2 = \mathrm{Var}_\pi(H_t f)$. In terms of u, the Nash's inequality above gives

$$\forall\, t,\ \ u(t)^{1+2/d} \le -\frac{C}{2}u'(t),$$

since $\|f\|_1 = 1$ implies $\|H_t f\|_1 \le 1$ for all $t > 0$. Setting $v(t) = \frac{dC}{4}u(t)^{-2/d}$ this differential inequality implies $v'(t) \ge 1$. Thus $v(t) \ge t$ (because $v(0) \ge 0$). Finally,

$$\forall\, t>0,\ \ u(t) \le \left(\frac{dC}{4t}\right)^{d/2}.$$

Taking the supremum over all functions f with $\|f\|_1 = 1$ yields

$$\forall\, t,\ \ \|H_t - \pi\|_{1\to 2} \le \left(\frac{dC}{4t}\right)^{d/4}.$$

The same applies to adjoint H_t^* and thus

$$\forall\, t>0,\ \ \|H_t - \pi\|_{2\to\infty} \le \left(\frac{dC}{4t}\right)^{d/4}.$$

Finally, using $H_t - \pi = (H_{t/2} - \pi)(H_{t/2} - \pi)$, we get

$$\forall\, t>0,\ \ \|H_t - \pi\|_{1\to\infty} \le \left(\frac{dC}{2t}\right)^{d/2}$$

which is the same as

$$|h_t(x,y) - 1| \le (dC/2t)^{d/2}.$$

Theorem 2.3.1 *Assume that the finite Markov chain (K,π) satisfies*

$$\forall\, g \in \ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}. \tag{2.3.1}$$

Then

$$\forall\, t>0,\ \ \|h_t^x - 1\|_2 \le \left(\frac{dC}{4t}\right)^{d/4}$$

and

$$\forall\, t>0,\ \ |h_t(x,y) - 1| \le \left(\frac{dC}{2t}\right)^{d/2}.$$

Let us discuss what this says. First, the hypothesis 2.3.1 and Jensen's inequality imply $\forall\ g \in \ell^2(\pi),\quad \mathrm{Var}_\pi(g) \leq C\mathcal{E}(g,g)$. This is a Poincaré inequality and it shows that $\lambda \geq 1/C$. Thus, the conclusion of Theorem 2.3.1 must be compared with

$$\forall\ t > 0,\ \ \|h_t^x - 1\|_2 \leq \pi(x)^{-1/2} e^{-t/C} \tag{2.3.2}$$

which follows from Corollary 2.1.5 when $\lambda \geq 1/C$. This last inequality looks better than the conclusion of Theorem 2.3.1 as it gives an exponential rate. However, Theorem 2.3.1 gives $\|h_t^x - 1\|_2 \leq 1$ for $t = dC/4$ whereas, for the same t, the right hand side of (2.3.2) is equal to $\pi(x)^{-1/2}e^{-d/4}$. Thus, if d is small and $1/\pi(x)$ large, the conclusion of Theorem 2.3.1 improves up on (2.3.2) at least for relatively small value of t. Assume for instance that (2.3.1) holds with $C = A/\lambda$ where we think of A as a numerical constant. Then, for $\theta = dA/(4\lambda)$, $\|H_\theta - \pi\|_{2\to\infty} = \max_x \|h_\theta^x - 1\|_2 \leq 1$. Hence, for $t = s + \theta = s + dA/(4\lambda)$

$$\begin{aligned} \|h_t^x - 1\|_2 &\leq \|(H_s - \pi)(H_\theta - \pi)\|_{2\to\infty} \\ &\leq \|H_s - \pi\|_{2\to 2}\|H_\theta - \pi\|_{2\to\infty} \\ &\leq e^{-\lambda s}. \end{aligned}$$

This yields

Corollary 2.3.2 *If* (K,π) *satisfies* (2.3.1) *with some constants* $C, d > 0$. *Then* $\lambda \geq 1/C$ *and*

$$\forall\ t > 0,\ \ \|h_t^x - 1\|_2 \leq \min\left\{ (dC/4t)^{d/4}, e^{-(t-\frac{dC}{4})\lambda} \right\}.$$

If (K,π) is reversible, then K is self-adjoint on $\ell^2(\pi)$ and $1-\lambda$ is the second largest eigenvalue of K. Consider an eigenfunction ψ for the eigenvalue $1-\lambda$, normalized so that $\max|\psi| = 1$. Then,

$$\begin{aligned} \max_x \|H_t^x - \pi\|_1 &= \max_{\|f\|_\infty \leq 1} \|(H_t - \pi)f\|_\infty \\ &\geq \|(H_t - \pi)\psi\|_\infty \\ &= e^{-t\lambda}. \end{aligned}$$

Hence

Corollary 2.3.3 *Assume that* (K,π) *is a reversible Markov chain. Then*

$$e^{-\lambda t} \leq \max_x \|H_t^x - \pi\|_1.$$

Furthermore, if (K,π) *satisfies* (2.3.1) *with* $C = A/\lambda$ *then*

$$e^{-\lambda t} \leq \max_x \|H_t^x - \pi\|_1 \leq 2\, e^{-\lambda t + \frac{dA}{4}}$$

for all $t > 0$.

This illustrates well the strength of Nash inequalities. They produce sharp results in certain circumstances where the time needed to reach stationarity is approximatively $1/\lambda$.

2.3.2 Nash's argument for finite Markov chains II

We now presents a second version of Nash's argument for finite Markov chains which turns out to be often easier to use than Theorem 2.3.1 and Corollary 2.3.2.

Theorem 2.3.4 *Assume that the finite Markov chain (K, π) satisfies*

$$\forall\, g \in \ell^2(\pi),\ \ \|g\|_2^{2(1+2/d)} \le C \left\{ \mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2 \right\} \|g\|_1^{4/d}. \tag{2.3.3}$$

Then

$$\forall\, t \le T,\ \ \|h_t^x\|_2 \le e \left(\frac{dC}{4t} \right)^{d/4}$$

and

$$\forall\, t \le T,\ \ h_t(x,y) \le e \left(\frac{dC}{2t} \right)^{d/2}.$$

The idea behind Theorem 2.3.4 is that Nash inequalities are most useful to capture the behavior of the chain for relatively small time, i.e., time smaller than T. In contrast with (2.3.1) the Nash inequality (2.3.3) implies no lower bound on the spectral gap. This is an advantage as it allows (2.3.3) to reflect the early behavior of the chain without taking into account the asymptotic behavior. This is well illustrated by two examples that will be treated later in these notes. Consider the natural chain on a square grid $\mathcal{G}_n$ of side length n and the natural chain on the n-dog $\mathcal{D}_n$ obtained by gluing together two copies of $\mathcal{G}_n$ at one of their corners. On one hand the spectral gap of $\mathcal{G}_n$ is of order $1/n^2$ whereas the spectral gap of $\mathcal{D}_n$ is of order $1/[n^2 \log n]$ (these facts will be proved later on). On the other hand, $\mathcal{G}_n$ and $\mathcal{D}_n$ both satisfy a Nash inequality of type (2.3.3) with C and T of order n^2. That is, the chains on $\mathcal{G}_n$ and $\mathcal{D}_n$ have similar behaviors for t less than n^2 whereas their asymptotic behavior as t goes to infinity are different. This is not surprising since the local structure of these two graphs are the same. For $\mathcal{D}_n$ a constant C of order $n^2 \log n$ is necessary for an inequality of type (2.3.1) to hold true.

PROOF OF THEOREM 2.3.4: Fix f satisfying $\|f\|_1 = 1$ and set

$$u(t) = e^{-2t/T}\|H_t f\|_2^2.$$

Then

$$u'(t) = -2e^{-2t/T}\left(\mathcal{E}(H_t f, H_t f) + \frac{1}{T}\|H_t f\|_2^2 \right).$$

Thus, Nash's argument yields

$$u(t) \le \left(\frac{dC}{4t} \right)^{d/2}$$

which implies

$$\|H_t\|_{1\to 2} \le e^{t/T}\left(\frac{dC}{4t}\right)^{d/4}.$$

The announced results follow since

$$\max_x \|h_t^x\|_2 = \|H_t^*\|_{1\to 2} \le e^{t/T}\left(\frac{dC}{4t}\right)^{d/4}$$

by the same argument applied to H_t^*.

Corollary 2.3.5 *Assume that (K,π) satisfies (2.3.3) and has spectral gap λ. Then for all $c \ge 0$ and all $0 < t_0 \le T$,*

$$\|h_t^x - 1\|_2 \le e^{1-c}$$

and

$$|h_{2t}(x,y) - 1| \le e^{2-2c}$$

for

$$t = t_0 + \frac{1}{\lambda}\left(\frac{d}{4}\log\left(\frac{dC}{4t_0}\right) + \ c\right).$$

PROOF: Write $t = s + t_0$ with $t_0 \le T$ and

$$\begin{aligned}\|h_t^x - 1\|_2 &\le \|(H_s - \pi)H_{t_0}\|_{2\to\infty}\\ &\le \|H_s - \pi\|_{2\to 2}\|H_{t_0}\|_{2\to\infty}\\ &\le e(dC/4t_0)^{d/4}\, e^{-\lambda s}.\end{aligned}$$

The result easily follows.

In practice, a "good" Nash inequality is (2.3.3) with a small value of d and $C \approx T$. Indeed, if (2.3.3) holds with, say $d = 4$ and $C = T$, then taking $t_0 = T$ in Corollary 2.3.5 yields

$$\|h_t^x - 1\|_2 \le e^{1-c} \text{ for } t = T + c/\lambda.$$

We now give a simple example that illustrates the strength of a good Nash inequality.

EXAMPLE 2.3.1: Consider the Markov chain on $\mathcal{X} = \{-n, \ldots, n\}$ with Kernel $K(x,y) = 0$ unless $|x - y| = 1$ or $x = y = \pm n$ in which cases $K(x,y) = 1/2$. This is an irreducible chain which is reversible with respect to $\pi \equiv (2n+1)^{-1}$. The Dirichlet form of this chain is given by

$$\mathcal{E}(f,f) = \frac{1}{2n+1}\sum_{-n}^{n-1}|f(i+1) - f(i)|^2.$$

For any $u, v \in \mathcal{X}$, and any function f, we have

$$|f(v) - f(u)| \le \sum_{i,i+1 \text{ between } u,v} |f(i+1) - f(i)|.$$

Hence, if f is not of constant sign,

$$\|f\|_\infty \le \sum_{-n}^{n-1} |f(i+1) - f(i)|.$$

To see this take u to be such that $\|f||_\infty = f(u)$ and v such that $f(v)f(u) \le 0$ so that $|f(u) - f(v)| \ge |f(u)|$. Fix a function g such that $\pi(g > 0) \le 1/2$ and $\pi(g < 0) \le 1/2$ (i.e., 0 is a median of g). Set $f = \mathrm{sgn}(g)|g|^2$. Then f changes sign. Observe also that

$$\begin{aligned} |f(i+1) - f(i)| &= |\mathrm{sgn}(g(i+1))g(i+1)^2 - \mathrm{sgn}(g(i))g(i)^2| \\ &\le |g(i+1) - g(i)|(|g(i+1)| + |g(i)|). \end{aligned}$$

Hence

$$\begin{aligned} \|f\|_\infty &\le \sum_{-n}^{n-1} |f(i+1) - f(i)| \\ &\le \sum_{-n}^{n-1} |g(i+1) - g(i)|(|g(i+1)| + |g(i)|) \\ &\le \left(\sum_{-n}^{n-1} |g(i+1) - g(i)|^2\right)^{1/2} \left(\sum_{-n}^{n-1} (|g(i+1)| + |g(i)|)^2\right)^{1/2} \\ &\le 2^{1/2}(2n+1)\mathcal{E}(g,g)^{1/2}\|g\|_2. \end{aligned}$$

That is

$$\|g\|_\infty^2 \le 2^{1/2}(2n+1)\mathcal{E}(g,g)^{1/2}\|g\|_2.$$

It follows that

$$\begin{aligned} \|g\|_2^4 &\le \|g\|_\infty^2 \|g\|_1^2 \\ &\le 2^{1/2}(2n+1)\mathcal{E}(g,g)^{1/2}\|g\|_2\|g\|_1. \end{aligned}$$

Hence for any g with median 0,

$$\|g\|_2^6 \le 2(2n+1)^2\mathcal{E}(g,g)\|g\|_1^4.$$

For any f with median c, we can apply the above to $g = f - c$ to get

$$\|f - c\|_2^6 \le 2(2n+1)^2\mathcal{E}(f,f)\|f - c\|_1^4 \le 2(2n+1)^2\mathcal{E}(f,f)\|f\|_1^4.$$

Hence

$$\forall\, f, \quad \mathrm{Var}_\pi(f)^3 \le 2(2n+1)^2\mathcal{E}(f,f)\|f\|_1^4.$$

This is a Nash inequality of type (2.3.1) with $C = 2(2n+1)^2$ and $d = 1$. It implies that

$$\lambda \ge \frac{1}{2(2n+1)^2}$$

and, by Theorem 2.3.1 and Corollary 2.3.2

$$\forall\, t > 0, \qquad \|h_t^x - 1\|_2 \le \left(\frac{(2n+1)^2}{2t}\right)^{1/4}$$

and

$$\forall\, c > 0, \quad \|h_t^x - 1\|_2 \le e^{-c} \quad \text{with } t = \frac{1}{2(2n+1)^2}(4+c).$$

The test function $f(i) = \mathrm{sgn}(i)|i|$ shows that

$$\lambda \le \frac{12}{(2n+1)^2}$$

(in fact $\lambda = 1 - \cos(\pi/(2n+1))$). By Corollary 2.3.3 it follows that

$$e^{-\frac{12\,t}{(2n+1)^2}} \le \max_{\mathcal{X}} \|h_t^x - 1\|_1 \le 2e^{-\frac{t}{2(2n+1)^2}+\frac{1}{4}}.$$

This shows that a time of order n^2 is necessary and sufficient for approximate equilibrium. This conclusion must be compare with

$$\|h_t^x - 1\|_1 \le \sqrt{2n+1}\, e^{-\frac{t}{2(2n+1)^2}}$$

which follows by using only the spectral gap estimate $\lambda \ge 1/(2(2n+1)^2)$ and Corollary 2.1.5. This last inequality only shows that a time of order $n^2 \log n$ is sufficient for approximate equilibrium.

2.3.3 Nash inequalities and the log-Sobolev constant

Thanks to Theorem 2.2.13 and Nash's argument it is possible to bound the log-Sobolev constant α in terms of a Nash inequality.

Theorem 2.3.6 *Let (K,π) be a finite reversible Markov chain.*

1. *Assume that (K,π) satisfies* (2.3.1)*, that is,*

$$\forall\, g \in \ell^2(\pi), \;\; \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

Then the log-Sobolev constant α of the chain is bounded below by

$$\alpha \ge \frac{2}{dC}.$$

2. *Assume instead that (K,π) satisfies* (2.3.3)*, that is,*

$$\forall\, g \in \ell^2(\pi), \;\; \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d},$$

and has spectral gap λ. Then the log-Sobolev constant α is bounded below by

$$\alpha \ge \frac{\lambda}{2\left[1 + \lambda t_0 + \frac{d}{4}\log\left(\frac{dC}{4t_0}\right)\right]}$$

for any $0 < t_0 \le T$.

PROOF: For the first statement, observe that Theorem 2.3.1 gives $\|H_t - \pi\|_{2\to\infty} \le 1$ for $t = dC/4$. Pluging this into Theorem 2.2.13 yields $\alpha \ge 2/(dC)$, as desired.

For the second inequality use Theorem 2.3.3 with $t = t_0 \le T$ and Theorem 2.2.13.

EXAMPLE 2.3.2: Consider the Markov chain of Example 2.3.1 on $\mathcal{X} = \{-n, \dots, n\}$ with Kernel $K(x, y) = 0$ unless $|x - y| = 1$ or $x = y = \pm n$ in which cases $K(x, y) = 1/2$. We have proved that it satisfies the Nash inequality

$$\forall f, \quad \mathrm{Var}_\pi(f)^3 \le 2(2n+1)^2 \mathcal{E}(f, f)\|f\|_1^4$$

of type (2.3.1) with $C = 2(2n+1)^2$ and $d = 1$. Hence Theorem 2.3.6 yields

$$\alpha \ge \frac{1}{(2n+1)^2}.$$

2.3.4 A converse to Nash's argument

Carlen et al. [11] found that there is a converse to Nash's argument. We now present a version of their result.

Theorem 2.3.7 *Assume that (K, π) is reversible and satisfies*

$$\forall\, t \le T, \quad \|H_t\|_{1\to 2} \le \left(\frac{C}{t}\right)^{d/4}.$$

Then

$$\forall\, g \in \ell^2(\pi), \quad \|f\|_2^{2(1+2/d)} \le C' \left(\mathcal{E}(f, f) + \frac{1}{2T}\|f\|_2^2\right)\|f\|_1^{4/d}$$

with $C' = 2^{2(1+2/d)}C$.

PROOF: Fix f with $\|f\|_1 = 1$ and write, for $0 < t \le T$,

$$\begin{aligned} \|f\|_2^2 &= \|H_t f\|_2^2 - \int_0^t \partial_s \|H_s f\|_2^2 ds \\ &= \|H_t f\|_2^2 + 2\int_0^t \mathcal{E}(H_s f, H_s f) ds \\ &\le (C/t)^{d/2} + 2t\mathcal{E}(f, f). \end{aligned}$$

The inequality uses the hypothesis (which implies $\|H_t f\|_2 \le (C/t)^{d/4}$ because $\|f\|_1 \le 1$) and the fact that $t \to \mathcal{E}(H_t f, H_t f)$ is nonincreasing, a fact that uses reversibility. This can be proved by writing

$$\mathcal{E}(H_t f, H_t f) = \|(I - K)^{1/2} H_t f\|_2^2 \le \|(I - K)^{1/2} f\|_2^2 = \mathcal{E}(f, f).$$

It follows that

$$\|f\|_2^2 \le (C/t)^{d/2} + 2t\left(\mathcal{E}(f, f) + \frac{1}{2T}\|f\|_2^2\right)$$

for all $t > 0$. The right-hand side is a minimum for

$$\frac{dC^{d/2}}{2} t^{-(1+d/2)} = 2\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)$$

and the minimum is

$$\left[(2/d)^{1/(1+2/d)} + (d/2)^{1/(1+d/2)}\right]\left[2C\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)\right]^{1/(1+2/d)}.$$

This yields

$$\|f\|_2^{2(1+2/d)} \leq B\left(\mathcal{E}(f,f) + \frac{1}{2T}\|f\|_2^2\right)$$

with

$$\begin{aligned} B &= 2C\left[(2/d)^{1/(1+2/d)} + (d/2)^{1-1/(1+2/d)}\right]^{1+2/d} \\ &= 2C(1+2/d)(1+d/2)^{2/d} \leq 2^{2+2/d}C. \end{aligned}$$

2.3.5 Nash inequalities and higher eigenvalues

We have seen that a Poincaré inequality is equivalent to a lower bound on the spectral gap λ (i.e., the smallest non-zero eigenvalue of $I-K$). It is interesting to note that Nash inequalities imply bounds on higher eigenvalues. Compare with [14].

Let (K,π) be a finite reversible Markov chain. Let $1 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{n-1}$ be the eigenvalues of $I-K$ and

$$N(s) = N_K(s) = \#\{i \in \{0,\ldots,n-1\} : \lambda_i \leq s\}, \quad s \geq 0,$$

be the eigenvalue counting function. Thus, N is a step function with $N(s) = 1$ for $0 \leq s < \lambda_1$ if (K,π) is irreducible. It is easy to relate the function N to the trace of the semigroup $H_t = e^{-t(I-K)}$. Since (K,π) is reversible, we have

$$\zeta(t) = \sum_x h_t(x,x)\pi(x) = \sum_x \|h^x_{t/2}\|_2^2 \pi(x) = \sum_{i=0}^{n-1} e^{-t\lambda_i}.$$

If $\lambda_i \leq 1/t$ then $e^{-t\lambda_i} \geq e^{-1}$. Hence

$$N(1/t) \leq e\zeta(t).$$

Now, it is clear that Theorems 2.3.1, 2.3.4 give upper bounds on ζ in terms of Nash inequalities.

Theorem 2.3.8 *Let (K,π) be a finite reversible Markov chain.*

1. *Assume that* (K,π) *satisfies* (2.3.1)*, that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

   Then the counting function N *satisfies*

$$N(s) \le 1 + e(dCs/2)^{d/2}$$

   for all $s \ge 0$.

2. *Assume instead that* (K,π) *satisfies* (2.3.3)*, that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d}.$$

   Then

$$N(s) \le e^3(dCs/2)^{d/2}$$

   for all $s \ge 1/T$.

Clearly, if $M(s)$ is a continuous increasing function such that $N(s) \le M(s)$, $s \ge 1/T$, then

$$\lambda_i = \max\{s : N(s) \le i\} \ge M^{-1}(i+1)$$

for all $i > M(1/T) - 1$. Hence, we obtain

Corollary 2.3.9 *Let* (K,π) *be a finite reversible Markov chain. Let* $1 = \lambda_0 \le \lambda_1 \le \ldots \le \lambda_{n-1}$ *be the eigenvalues of* $I - K$.

1. *Assume that* (K,π) *satisfies* (2.3.1)*, that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \mathrm{Var}_\pi(g)^{(1+2/d)} \le C\mathcal{E}(g,g)\|g\|_1^{4/d}.$$

   Then

$$\lambda_i \ge \frac{2i^{2/d}}{e^{2/d}dC}$$

   for all $i \in 1,\ldots,n-1$.

2. *Assume instead that* (K,π) *satisfies* (2.3.3)*, that is,*

$$\forall\, g \in \ell^2(\pi),\ \ \|g\|_2^{2(1+2/d)} \le C\left\{\mathcal{E}(g,g) + \frac{1}{T}\|g\|_2^2\right\}\|g\|_1^{4/d}.$$

   Then

$$\lambda_i \ge \frac{2(i+1)^{2/d}}{e^{6/d}dC}$$

   for all $i > e^3(dC/(2T))^{d/2} - 1$.

EXAMPLE 2.3.3: Assume that (K, π) is reversible, has spectral gap λ, and satisfies the Nash inequality (2.3.1) with $C = A/\lambda$ and some d, where we think of A as a numerical constant (e.g., $A = 100$) and d as fixed. Then, the corollary above says that

$$\lambda_i \geq c\lambda i^{2/d}$$

for all $0 \leq i \leq n-1$ with $c^{-1} = e^{2/d} dA$.

EXAMPLE 2.3.4: For the natural graph structure on $\mathcal{X} = \{-n, \ldots, n\}$, we have shown in Example 2.3.1 that the Nash inequality

$$\operatorname{Var}_\pi(f)^3 \leq 2(2n+1)^2 \mathcal{E}(f,f) \|f\|_1^4$$

holds. Corollary 2.3.9 gives

$$\lambda_j \geq \left(\frac{j}{e^2(2n+1)}\right)^2.$$

In this case, all the eigenvalues are known. They are given by

$$\lambda_j = 1 - \cos\frac{\pi j}{2n+1}, \quad 0 \leq j \leq 2n.$$

This compares well with our lower bound.

EXAMPLE 2.3.5: For a square grid on $\mathcal{X} = \{0, \ldots, n\}^2$, we will show later (Theorem 3.3.14) that

$$\operatorname{Var}_\pi(f)^2 \leq 64(n+1)^2 \mathcal{E}(f,f) \|f\|_1^2.$$

From this and corollary 2.3.9 we deduce

$$\lambda_i \geq \frac{i}{e2^7(n+1)^2}$$

for all $0 \leq i \leq (n+1)^2 - 1$. One can show that this lower bound is of the right order of magnitude for all i, n. Indeed the eigenvalues of this chain are the numbers

$$1 - \frac{1}{2}\left(\cos\frac{\pi\ell}{n+1} + \cos\frac{\pi k}{n+1}\right), \quad \ell, k \in \{0, \ldots, n\}$$

which are distributed roughly like

$$\frac{\ell^2 + k^2}{(n+1)^2}, \quad \ell, k \in \{0, \ldots, n\}$$

and we have

$$\#\left\{(\ell, k) \in \{0, \ldots, n\}^2 : \ell^2 + k^2 \leq j\right\} \simeq j.$$

2.3.6 Nash and Sobolev inequalities

Nash inequalities are closely related to the better known Sobolev inequalities (for some fixed $d > 2$)

$$\|f - \pi(f)\|_{2d/(d-2)}^2 \leq C\mathcal{E}(f,f), \tag{2.3.4}$$

$$\|f\|_{2d/(d-2)}^2 \leq C\left\{\mathcal{E}(f,f) + \frac{1}{T}\|f\|_2^2\right\}. \tag{2.3.5}$$

Indeed, the Hölder inequality

$$\|f\|_2^{2(1+2/d)} \leq \|f\|_{2d/(d-2)}^2 \|f\|_1^{4/d}$$

shows that the Sobolev inequality (2.3.4) (resp. (2.3.5)) implies the Nash inequality (2.3.1) (resp. (2.3.3)) with the same constants d, C, T. The converse is also true. (2.3.1) (resp. (2.3.3)) implies (2.3.4) (resp. (2.3.5)) with the same d, T and a C that differ only by a numerical multiplicative factor for large d. See [9].

We now give a complete argument showing that (2.3.1) implies (2.3.4), in the spirit of [9]. The same type of argument works for (2.3.3)) implies (2.3.5).

For any function $f \geq 0$ and any k, we set $f_k = (f - 2^k)_+ \wedge 2^k$ where $(t)_+ = \max\{0, t\}$ and $t \wedge s = \min\{t, s\}$. Thus, f_k has support in $\{x : f(x) > 2^k\}$, $f_k(x) = 2^k$ if $x \in \{z : f(z) \geq 2^{k+1}\}$ and $f_k = f - 2^k$ on $\{x : 2^k \leq f \leq 2^{k+1}\}$.

Lemma 2.3.10 *Let K be a finite Markov chain with stationary measure π. With the above notation, for any function f,*

$$\sum_k \mathcal{E}(|f|_k, |f|_k) \leq 2\mathcal{E}(f,f).$$

Proof: Since $\mathcal{E}(|f|, |f|) \leq \mathcal{E}(f,f)$, we can assume that $f \geq 0$. We can also assume that $K(x,y)\pi(x)$ is symmetric (if not use $\frac{1}{2}(K(x,y)\pi(x) + K(y,x)\pi(y))$). Observe that $|f_k(x) - f_k(y)| \leq |f(x) - f(y)|$ for all x, y. Write

$$\mathcal{E}(f_k, f_k) = \sum_{\substack{x,y \\ f(x) > f(y)}} (f_k(x) - f_k(y))^2 K(x,y)\pi(x).$$

Set

$$\begin{aligned} B_k &= \{x : 2^k < f(x) \leq 2^{k+1}\}, \\ B_k^- &= \{x : f(x) \leq 2^k\}, \\ B_k^+ &= \{x : 2^{k+1} < f(x)\}. \end{aligned}$$

Then

$$\mathcal{E}(f_k, f_k) =$$

$$
\begin{aligned}
& 2^{2k} \sum_{\substack{x \in B_k^+ \\ y \in B_k^-}} K(x,y)\pi(x) + \sum_{\substack{x \in B_k, y \in B_{k+1}^- \\ f(x) > f(y)}} (f_k(x) - f_k(y))^2 K(x,y)\pi(x) \\
\leq \; & 2^{2k} \sum_{\substack{x \in B_k^+ \\ y \in B_k^-}} K(x,y)\pi(x) + \sum_{\substack{x \in B_k, y \in \mathcal{X} \\ f(x) > f(y)}} (f(x) - f(y))^2 K(x,y)\pi(x) \\
= \; & A_1(k) + A_2(k).
\end{aligned}
$$

We now bound $\sum_k A_1(k)$ and $\sum_k A_2(k)$ separately.

$$\sum_k A_1(k) = \sum_{\substack{x,y \\ f(x) > f(y)}} \sum_{k : f(y) \leq 2^k < f(x)/2} 2^{2k}.$$

For x, y fixed, let k_0 be the smallest integer such that $f(y) \leq 2^{k_0}$ and k_1 be the largest integer such that $2^{k_1} < f(x)$. Then

$$\sum_{k : f(y) \leq 2^k < f(x)/2} 2^{2k} = \sum_{k=k_0}^{k_1 - 1} 4^k = \frac{1}{3}(4^{k_1} - 4^{k_0}) \leq (f(x) - f(y))^2.$$

The last inequality follows from the elementary inequality

$$a^2 - b^2 \leq 3(a-b)^2 \text{ if } a \geq 2b \geq 0.$$

This shows that

$$\sum_k A_1(k) \leq \mathcal{E}(f,f).$$

To finish the proof, note that

$$\sum_k A_2(k) = \sum_k \sum_{\substack{x \in B_k, y \in \mathcal{X} \\ f(x) > f(y)}} (f(x) - f(y))^2 K(x,y)\pi(x) = \mathcal{E}(f,f).$$

Lemma 2.3.10 is a crucial tool for the proof of the following theorem.

Theorem 2.3.11 *Assume that* (K, π) *satisfies the Nash inequality* (2.3.1)*, that is,*

$$\mathrm{Var}_\pi(g)^{(1+2/d)} \leq C\mathcal{E}(g,g)\|g\|_1^{4/d}$$

for some $d > 2$ *and all functions* g*. Then*

$$\|g - \pi(g)\|_{2d/(d-2)}^2 \leq B(d)C\mathcal{E}(g,g)$$

where $B(d) = 4^{6+2d/(d-2)}$.

PROOF: Fix a function g and let c denote a median of g. Consider the functions $f_\pm = (g-c)_\pm$ where $(t)_\pm = \max\{0, \pm t\}$. By definition of a median, we have

$$\pi\left(\{x : f_\pm(x) = 0\}\right) \geq 1/2.$$

For simplicity of notation, we set $f = f_+$ or f_-. For each k we define $f_k = (f - 2^k)_+ \wedge 2^k$ as in the proof of Lemma 2.3.10. Applying (2.3.1) to each f_k and setting $\pi_k = \pi(f_k)$, we obtain

$$\left[2^{2(k-1)}\pi(|f_k - \pi_k| \geq 2^{k-1})\right]^{1+2/d} \leq C\mathcal{E}(f_k, f_k)\left[2^k\pi(f \geq 2^k)\right]^{4/d}. \tag{2.3.6}$$

Observe that

$$\pi(\{x : f_k(x) = 0\}) \geq 1/2$$

and that, for any function $h \geq 0$ such that $\pi(\{x : h(x) = 0\}) \geq 1/2$ we have

$$\forall\, s \geq 0,\ \forall\, a,\ \ \pi(\{h \geq s\}) \leq 2\pi(\{|h - a| \geq s/2\}). \tag{2.3.7}$$

Indeed, if $a \leq s/2$ then $\pi(\{|h - a| \geq s/2\}) \geq \pi(h \geq s)$ whereas if $a \geq s/2$ then $\pi(\{|h - a| \geq s/2\}) \geq \pi(h = 0) \geq 1/2$. Using (2.3.6) and (2.3.7) with $h = f_k$, $a = \pi_k$ we obtain

$$\left[2^{2(k-1)}\pi(f_k \geq 2^k)\right]^{1+2/d} \leq 2^{1+2/d}C\mathcal{E}(f_k, f_k)\left[2^k\pi(f \geq 2^k)\right]^{4/d}.$$

Now, set $q = 2d/(d-2)$, $b_k = 2^{qk}\pi(\{f \geq 2^k\})$ and $\theta = d/(d+2)$. The last inequality (raised to the power θ) yields, after some algebra,

$$b_{k+1} \leq 2^{3+q}C^\theta\mathcal{E}(f_k, f_k)^\theta\, b_k^{2(1-\theta)}.$$

By Hölder's inequality

$$\begin{aligned}\sum_k b_k = \sum_k b_{k+1} &\leq 2^{3+q}C^\theta\left(\sum_k \mathcal{E}(f_k, f_k)\right)^\theta\left(\sum_k b_k^2\right)^{1-\theta} \\ &\leq 2^{3+q+\theta}C^\theta\mathcal{E}(f, f)^\theta\left(\sum_k b_k\right)^{2(1-\theta)}.\end{aligned}$$

It follows that

$$\left(\sum_k b_k\right)^{2\theta-1} \leq 2^{3+q+\theta}C^\theta\left(\sum_k \mathcal{E}(f_k, f_k)\right)^\theta.$$

Furthermore $2\theta - 1 = 2\theta/q$ and

$$\begin{aligned}(2^q - 1)\sum_k b_k &= \sum_k (2^{q(k+1)} - 2^{qk})\pi(\{f \geq 2^k\}) \\ &= \sum_k (2^{q(k+1)}\pi(\{2^k \leq g < 2^{k+1}\}) \geq \|f\|_q^q.\end{aligned}$$

Hence

$$\|f\|_q^2 \leq 2^{1+(3+q)/\theta}(2^q - 1)^{2/q}C\mathcal{E}(f, f).$$

Recall that $f = f_+$ or f_- with $f_\pm = (g-c)_\pm$, c a median of g. Note also that $\theta > 1/2$ when $d > 2$. Adding the inequalities for f_+ and f_- we obtain

$$\|g-c\|_q^2 \le 2(\|f_+\|_q^2 + \|f_-\|_q^2) \le 4^{5+q} C\mathcal{E}(g,g)$$

because $\mathcal{E}(f_+,f_+) + \mathcal{E}(f_-,f_-) \le \mathcal{E}(g,g)$. This easily implies that

$$\|g-\pi(g)\|_q^2 \le 4^{6+q} C\mathcal{E}(g,g)$$

which is the desired inequality. The constant 4^{6+q} can be improved by using a ρ-cutting, $\rho > 1$, instead of a dyadic cutting in the above argument. See [9].

2.4 Distances

This section discusses the issue of choosing a distance between probability distribution to study the convergence of finite Markov chains to their stationary measure. From the asymptotic point of view, this choice does not matter much. From a more quantitative point of view, it does matter sometimes but it often happen that different choices lead to similar results. This is a phenomenon which is not yet well understood. Many aspects of this question will not be considered here.

2.4.1 Notation and inequalities

Let μ, π be two probability measures on a finite set $\mathcal{X}$ (we work with a finite $\mathcal{X}$ but most of what is going to be said holds without any particlar assumption on $\mathcal{X}$). We consider π has the reference measure. Total variation is arguably the most natural distance between probability measures. It is defined by

$$\|\mu-\pi\|_{\mathrm{TV}} = \max_{A\subset\mathcal{X}} |\mu(A)-\pi(A)| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x)-\pi(x)|.$$

To see the second equality, use $\sum_x(\mu(x)-\pi(x)) = 0$. Note also that

$$\|\mu-\pi\|_{\mathrm{TV}} = \max\{|\mu(f)-\pi(f)| : |f| \le 1\}$$

where $\mu(f) = \sum_x f(x)\mu(x)$. A well known result in Markov chain theory relates total variation with the coupling technique. See, e.g., [4, 17] and the references therein.

All the others metrics or metric type quantities that we will consider are defined in terms of the density of μ with respect to π. Hence, set $h = \mu/\pi$. The ℓ^p distances

$$\|h-1\|_p = \left(\sum_{x\in\mathcal{X}} |h(x)-1|^p \pi(x)\right)^{1/p}, \quad \|h-1\|_\infty = \max_{x\in\mathcal{X}} |h(x)-1|$$

are natural choices for the analyst and will be used throughout these notes. The case $p = 2$ is of special interest as it brings in a useful Hilbert space structure.

It is known to statisticians as the chi-square distance. The case $p = 1$ is nothing else that total variation since

$$\|h - 1\|_1 = \sum_{x \in \mathcal{X}} |h(x) - 1|\pi(x) = \sum_{x \in \mathcal{X}} |\mu(x) - \pi(x)| = 2\|\mu - \pi\|_{\mathrm{TV}}.$$

Jensen's inequality yields a clear ordering between these distances since it implies

$$\|h - 1\|_r \le \|h - 1\|_s \quad \text{for all} \quad 1 \le r \le s \le \infty.$$

If we view (as we may) μ, π as linear functionals $\mu, \pi : \ell^p(\pi) \to \mathbb{R}$, $f \to \mu(f), \pi(f)$, then

$$\|\mu - \pi\|_{\ell^p(\pi) \to \mathbb{R}} = \sup \{|\mu(f) - \pi(f)| : \|f\|_p \le 1\} = \|h - 1\|_q$$

where q is given by $1/p + 1/q = 1$ (see also Section 1.3.1). Most of the quantitative results described in these notes are stated in terms of the ℓ^2 and ℓ^∞ distances.

There are at least three more quantities that appear in the literature. The Kullback-Leibler separation, or entropy, is defined by

$$\mathrm{Ent}_\pi(h) = \sum_{x \in \mathcal{X}} [h(x) \log h(x)] \pi(x).$$

Observe that $\mathrm{Ent}_\pi(h) \ge 0$ by Jensen inequality. The Hellinger distance is

$$\begin{aligned} \|\mu - \pi\|_H &= \sum_{x \in \mathcal{X}} \left| \sqrt{h(x)} - 1 \right|^2 \pi(x) = \sum_{x \in \mathcal{X}} \left| \sqrt{\mu(x)} - \sqrt{\pi(x)} \right|^2 \\ &= 2 \left(1 - \sum_{x \in \mathcal{X}} \sqrt{h(x)}\, \pi(x) \right). \end{aligned}$$

It is not obvious why this distance should be of particular interest. However, Kakutani proved the following. Consider an infinite sequence $(\mathcal{X}_i, \pi_i)$ of probability spaces each of which carries a second probability measure $\mu_i = h_i \pi_i$ which is absolutely continuous with respect to π_i. Let $\mathcal{X} = \prod_i \mathcal{X}_i$, $\mu = \prod_i \mu_i$, $\pi = \prod_i \pi_i$. Kakutani's theorem asserts that μ is absolutely continuous with respect to π if and only if the product $\prod_i \left(\int_{\mathcal{X}_i} \sqrt{h_i}\, d\pi_i \right)$ converges.

Finally Aldous and Diaconis [4] introduces the notion of separation distance

$$d_{\mathrm{sep}}(\mu, \pi) = \max_{x \in \mathcal{X}} \{1 - h(x)\}$$

in connection with strong stationary (or uniform) stopping times. See [4, 17, 19]. Observe the absence of absolute value in this definition.

The next lemma collects inequalities between the various distances introduced above. These inequalities are all well known except possibly for the strange looking lower bounds in (2.4.2) and (2.4.4). The only inequality that uses the fact that $\mathcal{X}$ is discrete and finite is the upper bound in (2.4.1).

Lemma 2.4.1 *Let π and $\mu = h\pi$ be two probability measures on a finite set $\mathcal{X}$.*

1. *Set $\pi_* = \min_{\mathcal{X}} \pi$. For $1 \le r \le s \le \infty$,*

$$\|h-1\|_r \le \|h-1\|_s \le \pi_*^{1/s-1/r}\|h-1\|_r. \tag{2.4.1}$$

Also

$$\left(\|h-1\|_2^2 - \|h-1\|_3^3\right) \le \|h-1\|_1 \le \|h-1\|_2. \tag{2.4.2}$$

2. *The Hellinger distance satisfies*

$$\frac{1}{4}\|h-1\|_1^2 \le \|\mu-\pi\|_H \le \frac{1}{4}\|h-1\|_1 \tag{2.4.3}$$

and

$$\frac{1}{8}\left(\|h-1\|_2^2 - \|h-1\|_3^3\right) \le \|\mu-\pi\|_H \le \|h-1\|_2^2 \tag{2.4.4}$$

3. *The entropy satisfies*

$$\frac{1}{2}\|h-1\|_1^2 \le \mathrm{Ent}_\pi(h) \le \frac{1}{2}\left(\|h-1\|_1 + \|h-1\|_2^2\right). \tag{2.4.5}$$

4. *The separation $d_{\mathrm{sep}}(\mu,\pi)$ satisfies*

$$\frac{1}{2}\|h-1\|_1 \le d_{\mathrm{sep}}(\mu,\pi) \le \|h-1\|_\infty. \tag{2.4.6}$$

PROOF: The inequalities in (2.4.1) are well known (the first follows from Jensen's inequality). The inequalities in (2.4.6) are elementary.

The upper bound in (2.4.5) uses

$$\forall\, u>0,\quad (1+u)\log(1+u) \le u + \frac{1}{2}u^2$$

to bound the positive part of the entropy. The lower bound is more tricky. First, observe that

$$\forall\, u>0,\quad 3(u-1)^2 \le (4+2u)(u\log(u)-u+1).$$

Then take square roots and use Cauchy-Schwarz to obtain

$$3\|h-1\|_1^2 \le \|4+2h\|_1\ \|h\log(h)-h+1\|_1.$$

Finally observe that $u\log(u)-u+1 \ge 0$ for $u \ge 0$. Hence $\|h\log(h)-h+1\|_1 = \mathrm{Ent}_\pi(f)$ and

$$3\|h-1\|_1^2 \le 6\mathrm{Ent}_\pi(f)$$

which gives the desired inequality. In his Ph. D. thesis, F. Su noticed the complementary bound

$$\mathrm{Ent}_\pi(h) \le \log\left(1+\|h-1\|_2^2\right).$$

The upper bound in (2.4.3) follows from $|\sqrt{u}-1|^2 \le |\sqrt{u}-1|(\sqrt{u}+1) = |u-1|$, $u \ge 0$. The lower bound in (2.4.3) uses $|u-1| = |\sqrt{u}-1|(\sqrt{u}+1)$, $u \ge 0$, Cauchy-Schwarz, and $\|\sqrt{h}+1\|_2^2 \le 4$.

The upper bound in (2.4.4) follows from $|\sqrt{u}-1| \le |u-1|$, $u \ge 0$. For the lower bound note that

$$\sqrt{1+u} \le \begin{cases} 1+\frac{1}{2}u - \frac{1}{16}u^2 & \text{for } -1 \le u \le 1 \\ 1+\frac{1}{2}u \le 1+\frac{1}{2}u - \frac{1}{16}u^2 + \frac{1}{16}u^3 & \text{for } 1 \le u. \end{cases}$$

It follows that

$$\forall\,, u \ge -1, \quad \sqrt{1+u} \le 1 + \frac{1}{2}u - \frac{1}{16}u^2 + \frac{1}{16}|u|^3.$$

Now, $\|\mu - \pi\|_H = 2(1 - \|\sqrt{h}\|_1) = 2(1 - \|\sqrt{1+(h-1)}\|_1)$. Hence

$$\|\mu - \pi\|_H \ge \frac{1}{8}(\|h-1\|_2^2 - \|h-1\|_3^3)).$$

Finally, the upper bound in (2.4.2) is a special case of (2.4.1). The lower bound follows from the elementary inequality: $\forall\, u \ge -1, \quad |u| \ge \frac{3}{4}u + u^2 - |u|^3$. This ends the proof of Lemma 2.4.1.

2.4.2 The cutoff phenomenon and related questions

This Section describe briefly a surprising property appearing in number of examples of natural finite Markov chains where a careful study is possible. We refer the reader to [4, 17] and the more recent [18] for further details and references.

Consider the following example of finite Markov chain. The state space $\mathcal{X} = \{0,1\}^n$ is the set of all binary vectors of length n. At each step, we pick a coordinate at random and flip it to its opposite. Hence, the kernel K of the chain is $K(x,y) = 0$ unless $|x-y| = 1$ in which case $K(x,y) = 1/n$. This chain is symmetric, irreducible but periodic. It has the uniform distribution $\pi \equiv 2^{-n}$ as stationary measure. Let $H_t = e^{-t}\sum_0^\infty \frac{t^i}{i!}K^i$ be the associated continuous time chain. Then, by the Perron-Frobenius theorem $H_t(x,y) \to 2^{-n}$ as t tends to infinity. This can be quantified very precisely.

Theorem 2.4.2 *For the continuous time chain on the hypercube $\{0,1\}^n$ described above, let $t_n = \frac{1}{4}n\log n$. Then for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \|H^x_{(1-\varepsilon)t_n} - 2^{-n}\|_{\mathrm{TV}} = 1$$

whereas

$$\lim_{n\to\infty} \|H^x_{(1+\varepsilon)t_n} - 2^{-n}\|_{\mathrm{TV}} = 0$$

In fact, a more precise description is feasible in this case. See [20, 18]. This theorem exhibits a typical case of the so called *cutoff phenomenon.* For n large enough, the graph of $t \to y(t) = \|H^x_t - 2^{-n}\|_{\mathrm{TV}}$ stays very close to the line $y = 1$ for a long time, namely for about $t_n = \frac{1}{4}n\log n$. Then, it falls off rapidly to a value close to 0. This fall-off phase is much shorter than t_n. Reference [20] describes the shape of the curve around the critical time t_n.

Definition 2.4.3 *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain.*

1. *One says that $\mathcal{F}$ presents a cutoff in total variation with critical time $(t_n)_1^\infty$ if $t_n \to \infty$ and*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|H^x_{n,(1-\varepsilon)t_n} - \pi_n\|_{\mathrm{TV}} = 1$$
and
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|H^x_{n,(1+\varepsilon)t_n} - \pi_n\|_{\mathrm{TV}} = 0.$$

2. *Let $(t_n, b_n)_1^\infty$ such that $t_n, b_n \geq 0$, $t_n \to \infty$, $b_n/t_n \to 0$. One says that $\mathcal{F}$ presents a cutoff of type $(t_n, b_n)_1^\infty$ in total variation if for all real c*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|H^x_{n,t_n+b_nc} - \pi_n\|_{\mathrm{TV}} = f(c)$$
with $f(c) \to 1$ when $c \to -\infty$ and $f(c) \to 0$ when $c \to \infty$.

Clearly, $2 \Rightarrow 1$. The ultimate cutoff result consists in a precise description of the function f. In Theorem 2.4.2 there is in fact a (t_n, b_m)-cutoff with $t_n = \frac{1}{4} n \log n$ and $b_n = n$. See [20].

In practical terms, the cutoff phenomenon means the following: in order to approximate the stationary distribution π_n one should not stop the chain $H_{n,t}$ before $t = t_n$ and it is essentially useless to run the chain for more than t_n. It seems that the cutoff phenomenon is widespread among natural examples. See [4, 18]. Nevertheless it is rather difficult to verify that a given family of chains satisfy one or the other of the above two definitions. This motivates the following weaker definition.

Definition 2.4.4 *Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Fix $1 \leq p \leq \infty$.*

1. *One says that $\mathcal{F}$ presents a weak ℓ^p-cutoff with critical time $(t_n)_1^\infty$ if $t_n \to \infty$ and*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,t_n} - 1\|_{\ell^p(\pi_n)} > 0 \quad \textit{and} \quad \lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,(1+\varepsilon)t_n} - 1\|_{\ell^p(\pi_n)} = 0.$$

2. *Let $(t_n, b_n)_1^\infty$ such that $t_n, b_n \geq 0$, $t_n \to \infty$, $b_n/t_n \to 0$. One says that $\mathcal{F}$ presents a weak ℓ^p-cutoff of type $(t_n, b_n)_1^\infty$ if for all $c \geq 0$,*
$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,t_n+cb_n} - 1\|_{\ell^p(\pi_n)} = f(c)$$
with $f(0) > 0$ and $f(c) \to 0$ when $c \to \infty$.

The notion of weak cutoff extends readily to Hellinger distance or entropy. The advantage of this definition is that it captures some of the spirit of the cutoff phenomenon without requiring a too precise understanding of what happens at relatively small times.

Observe that a cutoff of type $(t_n, b_n)_1^\infty$ is equivalent to a cutoff of type $(t_n, ab_n)_1^\infty$ with $a > 0$ but that t_n can not always be replaced by s_n even if $t_n \sim s_n$.

Note also that if $(t_n)_1^\infty$ and $(s_n)_1^\infty$ are critical times for a family $\mathcal{F}$ (the same for t_n and s_n) then $\lim_{n\to\infty} t_n/s_n = 1$. Indeed, for any $\epsilon > 0$, we must have $(1+\epsilon)t_n > s_n$ and $(1+\epsilon)s_n > t_n$ for n large enough.

Definition 2.4.5 *Let (K, π) be a finite irreducible Markov chain. For $1 \le p \le \infty$ and $\varepsilon > 0$, define the parameter $T_p(K,\varepsilon) = T_p(\varepsilon)$ by*

$$T_p(\varepsilon) = \inf\{t > 0 : \max_x \|h_t^x - 1\|_p \le \varepsilon\}$$

where $H_t = e^{-t(I-K)}$ is the associated continuous time chain.

The next lemma shows that for reversible chains and $1 < p \le \infty$ the different T_p's cannot be too different.

Lemma 2.4.6 *Let (K, π) be a finite irreducible reversible Markov chain. Then, for $2 \le p \le +\infty$ and $\varepsilon > 0$, we have*

$$T_2(K,\varepsilon) \le T_p(K,\varepsilon) \le T_\infty(K,\varepsilon) \le 2\,T_2(K,\varepsilon^{1/2}).$$

Furthermore, for $1 < p \le 2$ and $m_p = 1 + \lceil (2-p)/[2(p-1)] \rceil$,

$$T_p(K,\varepsilon) \le T_2(K,\varepsilon) \le m_p\,T_p(K,\varepsilon^{1/m_p}).$$

PROOF: The first assertion is easy and left as an exercise. For the second we need to use the fact that

$$\max_x \|h_{u+v}^x - 1\|_q \le \left(\max_x \|h_u^x - 1\|_r\right)\left(\max_x \|h_v^x - 1\|_s\right) \tag{2.4.7}$$

for all $u, v > 0$ and $1 \le q, r, s \le +\infty$ related by $1+1/q = 1/r+1/s$. Fix $1 < p < 2$ and an integer j. Set, for $i = 1, \ldots, j-1$, $p_1 = p$, $1 + 1/p_{i+1} = 1/p_i + 1/p$, and $u_i = it/j$, $v_i = t/j$. Applying (2.4.7) $j-1$ times with $q = p_{i+1}$, $r = p_i$, $s = p$, $u = u_i$, $v = v_j$, we get

$$\max_{\mathcal{X}} \|h_t^x - 1\|_{p_j} \le \left(\max_{\mathcal{X}} \|h_{t/j}^x - 1\|_p\right)^j.$$

Now, $p_j = 1/p - (j-1)(1-1/p)$. Thus $p_j \ge 2$ for

$$j \ge 1 + (2-p)/[2(p-1)].$$

The desired result follows.

Theorem 2.4.7 *Fix $1 < p < \infty$ and $\varepsilon > 0$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Let λ_n be the spectral gap of K_n and set $t_n = T_p(K_n, \varepsilon)$. Assume that*

$$\lim_{n\to\infty} \lambda_n t_n = \infty.$$

Then the family $\mathcal{F}$ presents a weak ℓ^p-cutoff of type $(t_n, 1/\lambda_n)_1^\infty$.

PROOF: By definition $\max_{\mathcal{X}_n} \|h^x_{n,t_n} - 1\|_p = \varepsilon > 0$. To obtain an upper bound write

$$\begin{aligned}
\|h^x_{n,t_n+s} - 1\|_p &= \|(H^*_{n,s} - \pi_n)(h^x_{n,t_n} - 1)\|_p \\
&\le \|h^x_{n,t_n} - 1\|_p \, \|H^*_{n,s} - \pi_n\|_{p\to p} \\
&\le \varepsilon \, \|H^*_{n,s} - \pi_n\|_{p\to p}.
\end{aligned}$$

By Theorem 2.1.4

$$\|H^*_{n,s} - \pi_n\|_{2\to 2} \le e^{-s\lambda_n}.$$

Also, $\|H^*_{n,s} - \pi_n\|_{1\to 1} \le 2$ and $\|H^*_{n,s} - \pi_n\|_{\infty\to\infty} \le 2$. Hence, by interpolation, (see Theorem 1.3.1)

$$\|H^*_{n,s} - \pi_n\|_{p\to p} \le 4^{|1/2-1/p|} e^{-s\lambda_n(1-2|1/2-1/p|)}.$$

It follows that

$$\|h^x_{n,t_n+c/\lambda_n} - 1\|_p \le \varepsilon 4^{|1/2-1/p|} e^{-c(1-2|1/2-1/p|)}.$$

This proves the desired result since $1 - 2|1/2 - 1/p| > 0$ when $1 < p < \infty$. This also proves the following auxilliary result.

Lemma 2.4.8 *Fix $1 < p < \infty$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of finite chains. Let λ_n be the spectral gap of K_n. If*

$$\lim_{n\to\infty} \lambda_n T_p(K_n, \varepsilon) \to \infty$$

for some fixed $\varepsilon > 0$, then

$$\lim_{n\to\infty} \frac{T_p(K_n, \varepsilon)}{T_p(K_n, \eta)} = 1.$$

for all $\eta > 0$.

For reversible chain we obtain a necessary and sufficient condition for weak ℓ^2-cutoff.

Theorem 2.4.9 *Fix $\varepsilon > 0$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of reversible finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Let λ_n be the spectral gap of K_n and set $t_n = T_2(K_n, \varepsilon)$. A necessary and sufficient condition for $\mathcal{F}$ to present a weak ℓ^2-cutoff with critical time t_n is that*

$$\lim_{n\to\infty} \lambda_n t_n = \infty. \tag{2.4.8}$$

Furthermore, if (2.4.8) *is satisfied then*

1. *$\mathcal{F}$ presents a weak ℓ^∞-cutoff of type $(2t_n, 1/\lambda_n)_1^\infty$.*

2. *For each $1 < p \leq \infty$ and each $\eta > 0$, $\mathcal{F}$ presents a weak ℓ^p-cutoff of type $(T_p(K_n, \eta), 1/\lambda_n)_1^\infty$.*

PROOF: We already now that (2.4.8) is sufficient to have a weak ℓ^2-cutoff. Conversely, if (2.4.8) does not hold there exists $a > 0$ and a subsequence $n(i)$ such that $\lambda_{n(i)} t_{n(i)} \leq a$. To simplify notation assume that this hold for all n. Let ϕ_n be an eigenfunction of K_n such that $\|\phi_n\|_\infty = 1$ and $(I - K_n)\phi_n = \lambda_n \phi_n$. Then

$$\max_{\mathcal{X}_n} \|h^x_{n,t} - 1\|_2 \geq \|(H^x_{n,t} - \pi_n)\phi_n\|_2 = e^{-t\lambda_n}.$$

If follows that, for any $\eta > 0$,

$$\max_{\mathcal{X}_n} \|h^x_{n,(1+\eta)t_n} - 1\|_2 \geq e^{-(1+\eta)t_n\lambda_n} \geq e^{-(1+\eta)a}.$$

Hence

$$\lim_{n\to\infty} \max_{\mathcal{X}_n} \|h^x_{n,(1+\eta)t_n} - 1\|_2 \not\to 0$$

which shows that there is no weak ℓ^2-cutoff.

To prove the assertion concerning the weak ℓ^∞-cutoff simply observe that

$$\max_{\mathcal{X}_n} \|h^x_{n,t} - 1\|_\infty = \max_{\mathcal{X}_n} \|h^x_{n,t/2} - 1\|_2^2.$$

Hence a weak ℓ^2-cutoff of type $(t_n, b_n)_1^\infty$ is equivalent to a weak ℓ^∞-cutoff of type $(2t_n, b_n)$.

For the last assertion use Lemmas 2.4.6 and 2.4.8 to see that (2.4.8) implies $\lambda_n T_p(K, \eta) \to \infty$ for any fixed $\eta > 0$. Then apply Theorem 2.4.7.

The following theorem is based on strong hypotheses that are difficult to check. Nevertheless, it sheds some new light on the cutoff phenomenon.

Theorem 2.4.10 *Fix $\varepsilon > 0$. Let $\mathcal{F} = \{(\mathcal{X}_n, K_n, \pi_n) : n = 1, 2, \ldots\}$ be an infinite family of reversible finite chains. Let $H_{n,t} = e^{-t(I-K_n)}$ be the corresponding continuous time chain. Let λ_n be the spectral gap of K_n and set $t_n = T_2(K_n, \varepsilon)$. Let α_n be the log-Sobolev constant of (K_n, π_n). Set*

$$A_n = \max\{\|\phi\|_\infty : \|\phi\|_2 = 1, K_n\phi = (1 - \lambda_n)\phi\}.$$

Assume that the following conditions are satisfied.

(1) $t_n \lambda_n \to \infty$.

(2) $\inf_n \{\alpha_n / \lambda_n\} = c_1 > 0$.

(3) $\inf_n \left\{A_n e^{-\lambda_n t_n}\right\} = c_2 > 0$.

Then the family $\mathcal{F}$ presents a weak ℓ^p-cutoff with critical time $(t_n)_1^\infty$ for any $1 \leq p < \infty$ and also in Hellinger distance.

PROOF: By Theorem 2.4.9 condition (1) implies a weak ℓ^p-cutoff of type

$$(T_p(K_n,\eta),\lambda_n)$$

for each $1<p<\infty$ and $\eta>0$. The novelty in Theorem 2.4.10 is that it covers the case $p=1$ (and Hellinger distance) and that the critical time $(t_n)_0^\infty$ **does not** depend on $1\le p<\infty$. For the case $p>2$, it suffices to prove that $T_p(K_n,\varepsilon)\le t_n+c(p)/\lambda_n$. Using symmetry, (2.2.2) and hypothesis (2), we get

$$\|h^x_{n,t_n+s_n}-1\|_p \le \|H_{n,s_n}\|_{2\to p}\|h^x_{n,t_n}-1\|_2\le\varepsilon$$

with $s_n=[\log(p-1)]/(4\alpha_n)\le[\log(p-1)]/(4c_1\lambda_n)$, which yields the desired inequality. Observe that condition (3) has not been used to treat the case $2<p<\infty$.

We now turn to the proof of the weak ℓ^1-cutoff. Since

$$\|h_{n,t}-1\|_1\le\|h_{n,t}-1\|_2$$

it suffices to prove that

$$\liminf_{n\to\infty}\|h_{n,t_n}-1\|_1>0.$$

To prove this, we use the lower bound in (2.4.2) and condition (3) above. Indeed, for each n there exists a normalized eigenfunction ϕ_n and $x_n\in\mathcal{X}_n$ such that $K_n\phi_n=(1-\lambda_n)\phi_n$ and $\|\phi_n\|_\infty=\phi_n(x_n)=A_n$. It follows that

$$\begin{aligned}\|h^{x_n}_{n,t_n+s}-1\|_2 &= \sup_{\|\psi\|_2\le1}\{\|(H_{n,t_n+s}-\pi_n)\psi\|_\infty\}\\ &\ge A_n\,e^{-\lambda_n(t_n+s)}\ge c_2\,e^{-\lambda_n s}.\end{aligned}$$

Also, for $\sigma_n=(\log 2)/(4\alpha_n)$, we have

$$\begin{aligned}\|h^x_{n,t_n+\sigma_n+s}-1\|_3 &\le \|h^x_{n,t_n+s}-1\psi\|_2\\ &\le \|h^x_{n,t_n}-1\psi\|_2\|H_{n,s}-\pi_n\|_{2\to2}\\ &\le \varepsilon\,e^{-\lambda_n s}.\end{aligned}$$

Hence, since $\lambda_n\sigma_n\le[\log 2]/4c_1$,

$$\begin{aligned}\|h^{x_n}_{n,t_n+\sigma_n+s}-1\|_1 &\ge \|h^{x_n}_{n,t_n+\sigma_n+s}-1\|_2^2-\|h^{x_n}_{n,t_n+\sigma_n+s}-1\|_3^3\\ &\ge c_2^2\,e^{-2\lambda_n(\sigma_n+s)}-\varepsilon^3\,e^{-3\lambda_n s}\\ &\ge \left(c_2^2\,e^{-2\lambda_n\sigma_n}-\varepsilon^3\,e^{-\lambda_n s}\right)\,e^{-2\lambda_n s}\\ &\ge \left(c_3-\varepsilon^3\,e^{-\lambda_n s}\right)\,e^{-2\lambda_n s}\end{aligned}$$

where $c_3=c_2^2 2^{-1/4c_1}$. For each fixed n, we now pick $s=s_n=\lambda_n^{-1}\log(c_3/(2\varepsilon^3))$. Hence

$$\|h^{x_n}_{n,t_n}-1\|_1\ge\|h^{x_n}_{n,t_n+\sigma_n+s_n}-1\|_1\ge c_3/2.$$

The weak cutoff in Hellinger distance is proved the same way using (2.4.3) or (2.4.4). Finally the case $1<p<2$ follows from the results obtained for $p=2$ and $p=1$.

Chapter 3

Geometric tools

This chapter uses adapted graph structures to study finite Markov chains. It shows how paths on graphs and their combinatorics can be used to prove Poincaré and Nash inequalities. Isoperimetric techniques are also considered. Path techniques have been introduced by M. Jerrum and A. Sinclair in their study of a stochastic algorithm that counts perfect matchings in a graph. See [72]. Paths are also used in [79] in a somewhat different context (random walk on finitely generated groups). They are used in [35] to prove Poincaré inequalities. The underlying idea is classical in analysis and geometry. The simplest instance of it is the following proof of a Poincaré inequality for the unit interval $[0,1]$:

$$\int_0^1 |f(s)-m|^2 ds \leq \frac{1}{8}\int_0^1 |f'(s)|^2 ds$$

where m is the mean of f. Write $f(s)-f(t)=\int_t^s f'(u)du$ for any $0\leq t<s\leq 1$. Hence, using the Cauchy-Schwarz inequality, $|f(s)-f(t)|^2 \leq (s-t)\int_t^s |f'(u)|^2 du$. It follows that

$$\begin{aligned}
\int_0^1 |f(s)-m|^2 ds &= \int_0^1\int_0^1 |f(s)-f(t)|^2 dt ds \\
&\leq \int_0^1 |f'(u)|^2 \left\{\int_0^1\int_0^1 (s-t)\mathbf{1}_{t\leq u\leq s}(u) dt ds\right\} du \\
&= \int_0^1 |f'(u)|^2 \left\{\frac{u(1-u)}{2}\right\} du \\
&\leq \frac{1}{8}\int_0^1 |f'(u)|^2 du.
\end{aligned}$$

The constant $1/8$ obtained by this argument must be compared with the best possible constant which is $1/\pi^2$.

This chapter develops and illustrates several versions of this technique in the context of finite graphs.

3.1 Adapted edge sets

Definition 3.1.1 *Let K be an irreducible Markov chain on a finite set $\mathcal{X}$. An edge set $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ is say to be adapted to K if $\mathcal{A}$ is symmetric (that is $(x, y) \in \mathcal{A} \Rightarrow (y, x) \in \mathcal{A}$), $(\mathcal{X}, \mathcal{A})$ is connected, and*

$$(x, y) \in \mathcal{A} \Rightarrow K(x, y) + K(y, x) > 0.$$

In this case we also say that the graph $(\mathcal{X}, \mathcal{A})$ is adapted.

Let K be an irreducible Markov kernel on $\mathcal{X}$ with stationary measure π. It is convenient to introduce the following notation. For any $e = (x, y) \in \mathcal{X} \times \mathcal{X}$, set

$$df(e) = f(y) - f(x)$$

and define

$$Q(e) = \frac{1}{2}\left(K(x, y)\pi(x) + K(y, x)\pi(y)\right).$$

We will sometimes view Q as a probability measure on $\mathcal{X} \times \mathcal{X}$. Observe that, by Definition 2.1.1 and (2.1.1), the Dirichlet form $\mathcal{E}$ of (K, π) satisfies

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{e \in \mathcal{X} \times \mathcal{X}} |df(e)|^2 Q(e).$$

Let $\mathcal{A}$ be an adapted edge set. A path γ in $(\mathcal{X}, \mathcal{A})$ is a sequence of vertices $\gamma = (x_0, \ldots, x_k)$ such that $(x_{i-1}, x_i) \in \mathcal{A}$, $i = 1, \ldots, k$. Equivalently, γ can be viewed as a sequence of edges $\gamma = (e_1, \ldots, e_k)$ with $e_i = (x_{i-1}, x_i) \in \mathcal{A}$, $i = 1, \ldots, k$. The length of such a path γ is $|\gamma| = k$. Let Γ be the set of all paths γ in $(\mathcal{X}, \mathcal{A})$ which have no repeated edges (that is, such that $e_i \neq e_j$ if $i \neq j$). For each pair $(x, y) \in \mathcal{X} \times \mathcal{X}$, set

$$\Gamma(x, y) = \{\gamma = (x_0, \ldots, x_k) \in \Gamma : x = x_0,\ y = x_k\}.$$

3.2 Poincaré inequality

A Poincaré inequality is an inequality of the type

$$\forall f, \quad \mathrm{Var}_\pi(f) \leq C\mathcal{E}(f, f).$$

It follows from the definition 2.1.3 of the spectral gap λ that such an inequality is equivalent to $\lambda \geq 1/C$. In other words, the smallest constant C for which the Poincaré inequality above holds is $1/\lambda$. This section uses Poincaré inequality and path combinatorics to bound λ from below. We start with the simplest result of this type.

Theorem 3.2.1 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set. For each $(x,y) \in \mathcal{X} \times \mathcal{X}$ choose exactly one path $\gamma(x,y)$ in $\Gamma(x,y)$. Then $\lambda \geq 1/A$ where*

$$A = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y \in \mathcal{X}: \\ \gamma(x,y) \ni e}} |\gamma(x,y)| \pi(x)\pi(y) \right\}.$$

PROOF: For each $(x,y) \in \mathcal{X} \times \mathcal{X}$, write

$$f(y) - f(x) = \sum_{e \in \gamma(x,y)} df(e)$$

and, using Cauchy-Schwarz,

$$|f(y) - f(x)|^2 \leq |\gamma(x,y)| \sum_{e \in \gamma(x,y)} |df(e)|^2.$$

Multiply by $\frac{1}{2}\pi(x)\pi(y)$ and sum over all x, y to obtain

$$\frac{1}{2} \sum_{x,y} |f(y) - f(x)|^2 \pi(x)\pi(y) \leq \frac{1}{2} \sum_{x,y} |\gamma(x,y)| \sum_{e \in \gamma(x,y)} |df(e)|^2 \pi(x)\pi(y).$$

The left-hand side is equal to $\mathrm{Var}_\pi(f)$ whereas the right-hand side becomes

$$\frac{1}{2} \sum_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y: \\ \gamma(x,y) \ni e}} |\gamma(x,y)| \pi(x)\pi(y) \right\} |df(e)|^2 Q(e)$$

which is bounded by

$$\max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y: \\ \gamma(x,y) \ni e}} |\gamma(x,y)| \pi(x)\pi(y) \right\} \mathcal{E}(f,f).$$

This proves the Poincaré inequality

$$\forall f, \quad \mathrm{Var}_\pi(f) \leq A\mathcal{E}(f,f)$$

hence $\lambda \geq 1/A$.

EXAMPLE 3.2.1: Let $\mathcal{X} = \{0,1\}^n$, $\pi \equiv 2^{-n}$ and $K(x,y) = 0$ unless $|x-y| = 1$ in which case $K(x,y) = 1/n$. Consider the obvious adapted edge set $\mathcal{A} = \{(x,y) : |x-y| = 1\}$. To define a path $\gamma(x,y)$ from x to y, view x, y as binary vectors and change the coordinates of x one at a time from left to right to match the coordinates of y. These paths have length at most n. Since $1/Q(e) = n\,2^n$ we obtain in this case

$$\begin{aligned} A &\leq n^2\, 2^{-n} \max_{e \in \mathcal{A}} \left\{ \sum_{\substack{x,y: \\ \gamma(x,y) \ni e}} 1 \right\} \\ &= n^2\, 2^{-n} \max_{e \in \mathcal{A}} \#\{(x,y) : \gamma(x,y) \ni e\}. \end{aligned}$$

Hence every thing boils down to count, for each edge $e \in \mathcal{A}$, how many paths $\gamma(x, y)$ use that edge. Let $e = (u, v)$. Since $e \in \mathcal{A}$, there exists a unique i such that $u_i \neq v_i$. Furthermore, by construction, if $\gamma(x, y) \ni e$ we must have

$$\begin{aligned} x &= (x_1, \ldots, x_{i-1}, u_i, u_{i+1}, \ldots, u_n) \\ y &= (v_1, \ldots, v_{i-1}, v_i, y_{i+1}, \ldots, y_n). \end{aligned}$$

It follows that $i - 1$ coordinates of x and $n - i$ coordinates of y are unknown. That is, $\#\{(x, y) : \gamma(x, y) \ni e\} = 2^{n-1}$. Hence $A \leq n^2/2$ and Theorem 3.2.1 yields $\lambda \geq 2/n^2$. The right answer is $\lambda = 2/n$. The above computation is quite typical of what has to be done to use Theorem 3.2.1. Observe in particular the non trivial cancellation of the exponential factors.

EXAMPLE 3.2.2: Keep $\mathcal{X} = \{0, 1\}^n$ and consider the following moves: $x \to \tau(x)$ where $\tau(x)_i = x_{i-1}$ and $x \to \sigma(x)$ where $\sigma(x) = x + (1, 0, \ldots, 0)$. Let $K(x, y) = 1/2$ if $y = \tau(x)$ or $y = \sigma(x)$ and $K(x, y) = 0$ otherwise. This chain has $\pi \equiv 2^{-n}$ as stationary distribution. It is not reversible. Define $\gamma(x, y)$ as follows. Use τ to turn the coordinates around from right to left. Use σ to ajust x_i to y_i if necessary as it passes in position 1. These paths have length at most $2n$. Let $e = (u, v)$ be an edge, say $v = \sigma(u)$. Pick an integer j, $0 \leq j \leq n - 1$. Then, if we assume that τ as been used exactly j times before e, then $x_i = u_{i-j}$ for $j < i \leq n$, $y_i = v_{n-j+i}$ for $1 \leq i \leq j$ and $y_{j+1} = v_1$. Hence, there are 2^{n-1} ordered pair (x, y) such that $e \in \gamma(x, y)$ appears after exactly j uses of τ. Since there are n possible values of j, this shows that the constant A of Theorem 3.2.1 is bounded by $A \leq 4n^2$ and thus $\lambda \geq 1/(4n^2)$.

EXAMPLE 3.2.3: Let again $\mathcal{X} = \{0, 1\}^n$. Let τ, σ be as in the preceding example. Consider the chain with kernel $K(x, y) = 1/n$ if either $y = \tau^j(x)$ for some $0 \leq j \leq n - 1$ or $y = \sigma(x)$, and $K(x, y) = 0$ otherwise. This chain is reversible with respect to the uniform distribution. Without further idea, it seems difficult to do any thing much better than using the same paths and the same analysis as in the previous example. This yields $A \leq n^3$ and $\lambda \geq 1/n^3$. Clearly, a better analysis is desirable in this case because we have not taken advantage of all the moves at our disposal. A better bound will be obtained in Section 4.2.

EXAMPLE 3.2.4: It is instructive to work out what Theorem 3.2.1 says for simple random walk on a graph $(\mathcal{X}, \mathcal{A})$ where $\mathcal{A}$ is a symmetric set of oriented edges. Set $d(x) = \#\{y \in \mathcal{X} : (x, y) \in \mathcal{A}\}$ and recall that the simple random walk on $(\mathcal{X}, \mathcal{A})$ has kernel

$$K(x, y) = \begin{cases} 0 & \text{if } (x, y) \notin \mathcal{A} \\ 1/d(x) & \text{if } (x, y) \in \mathcal{A}. \end{cases}$$

This gives a reversible chain with respect to the measure $\pi(x) = d(x)/|\mathcal{A}|$. For each $(x, y) \in \mathcal{X}^2$ choose a path $\gamma(x, y)$ with no repeated edge. Set

$$d_* = \max_{x \in \mathcal{X}} d(x), \quad \gamma_* = \max_{x, y \in \mathcal{X}} |\gamma(x, y)|, \quad \eta_* = \max_{e \in \mathcal{A}} \#\{(x, y) \in \mathcal{X}^2 : \gamma(x, y) \ni e\}.$$

Then Theorem 3.2.1 gives $\lambda \geq 1/A$ with

$$A \leq \frac{d_*^2 \gamma_* \eta_*}{|\mathcal{A}|}.$$

The quantity η_* can be interpreted as a measure of bottle necks in the graph $(\mathcal{X}, \mathcal{A})$. The quantity γ_* as an obvious interpretation as an upper bound on the diameter of the graph.

We now turn to more sophisticated (but still useful) versions of Theorem 3.2.1.

Definition 3.2.2 *A weight function w is a positive function*

$$w : \mathcal{A} \to (0, \infty).$$

The w-length of a path γ in Γ is

$$|\gamma|_w = \sum_{e \in \gamma} \frac{1}{w(e)}.$$

Theorem 3.2.3 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set and w be a weight function. For each $(x,y) \in \mathcal{X} \times \mathcal{X}$ choose exactly one path $\gamma(x,y)$ in $\Gamma(x,y)$. Then $\lambda \geq 1/A(w)$ where*

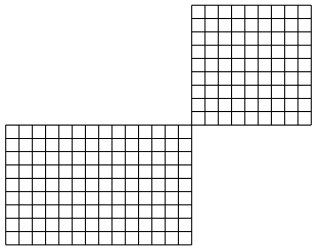
$$A(w) = \max_{e \in \mathcal{A}} \left\{ \frac{w(e)}{Q(e)} \sum_{\substack{(x,y):\\ \gamma(x,y) \ni e}} |\gamma(x,y)|_w \pi(x)\pi(y) \right\}.$$

Proof: Start as in the proof of Theorem 3.2.1 but introduce the weight w when using Cauchy-Schwarz to get

$$\begin{aligned} |f(y) - f(x)|^2 &\leq \left(\sum_{e \in \gamma(x,y)} w(e)^{-1} \right) \left(\sum_{e \in \gamma(x,y)} |df(e)|^2 w(e) \right) \\ &= |\gamma(x,y)|_w \sum_{e \in \gamma(x,y)} |df(e)|^2 w(e). \end{aligned}$$

From here, complete the proof by following step by step the proof of Theorem 3.2.1. A subtle discussion of this result can be found in [55] which also contains interesting examples.

Example 3.2.5: What is the spectral gap of the dog? (for simplicity, the dog below has no ears or legs or tail).

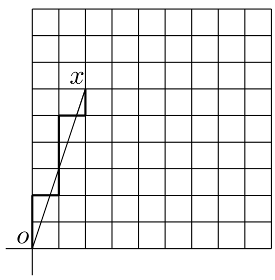


For a while, Diaconis and I puzzled over finding the order of magnitude of the spectral gap for simple random walk on the planar graph made from two square grids, say of side length n, attached together by one of their corners. This example became known to us as "the dog". It turns out that the dog is quite an interesting example. Thus, let $\mathcal{X}$ be the vertex set of two $n \times n$ square grids $\{0, \ldots, n\}^2$ and $\{-n, \ldots, 0\}^2$ attached by identifying the two corners $o = (0, 0) \in \mathcal{X}$ so that $|\mathcal{X}| = 2(n+1)^2 - 1$. Consider the markov kernel

$$K(x,y) = \begin{cases} 0 & \text{if } |x-y| > 1 \\ 1/4 & \text{if } |x-y| = 1 \\ 0 & \text{if } x = y \text{ is inside or } x = y = 0 \\ 1/4 & \text{if } x = y \text{ is on the boundary but not a corner} \\ 1/2 & \text{if } x = y \text{ is a corner.} \end{cases}$$

This is a symmetric kernel with uniform stationary measure $\pi \equiv (2(n+1)^2 - 1)^{-1}$ and $1/Q(e) = 4(2(n+1)^2 - 1)$ if $e \in \mathcal{A}$. We will refer to this example as the n-dog.

We now have to choose paths. The graph structure on $\mathcal{X}$ induces a distance $d(x, y)$ between vertices. Also, we have the Euclidean distance $|x - y|$. First we define paths from any $x \in \mathcal{X}$ to o. For definitness, we work in the square lying in the first quadrant. Let $\gamma(x, o)$ be one of the geodesic paths from x to o such that, for any $z \in \gamma(x, o)$, the Euclidean distance between z and the straight line segment $[x, o]$ is at most $1/\sqrt{2}$.



Let $e = (u, v)$ be an edge with $d(o, v) = i$, $d(o, u) = i + 1$. We claim that

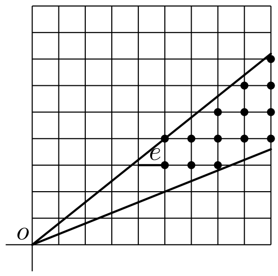
$$\#\{x : \gamma(x, o) \ni e\} \le \frac{4(n+1)^2}{i+1}.$$

By symmetry, we can assume that $u = (u_1, u_2)$ with $u_1 \geq u_2$. This implies that $u_1 \geq (i+1)/2$. Let I be the vertical segment of length 2 centred at u. Set

$$\{x : \gamma(x, o) \ni e\} = Z(e).$$

If $z \in Z(e)$ then the straight line segment $[o, z]$ is at Euclidean distance at most $1/\sqrt{2}$ from u. This implies that $Z(e)$ is contained in the half cone $\mathcal{C}(u)$ with vertex o and base I (because $(u_1 \geq u_2)$. Thus

$$Z(e) \subset \{(z_1, z_2) \in \{0, \ldots, n\}^2 : z_1 \geq u_1, z_2 \geq u_2\} \cap \mathcal{C}(u).$$



Let $\ell(j)$ be the length of the intersection of the vertical line $U(j)$ passing through $(j, 0)$ with $\mathcal{C}$. Then $\ell(j)/j = \ell(k)/k$ for all j, k. Clearly $\ell(u_1) = 3$. Hence $\ell(j) \leq 3j/u_1$. This means that there are at most $1 + 3j/u_1$ vertices in $U_j \cap Z(e)$. Summing over all $u_1 \leq j \leq n$ we obtain

$$\#Z(e) \leq n + \frac{3n(n+1)}{2u_1} \leq \frac{4n(n+1)}{i+1}.$$

which is the claimed inequality.

Now, if x, y are any two vertices in $\mathcal{X}$, we join them by going through o using the paths $\gamma(x, o), \gamma(y, o)$ in the obvious way. This defines $\gamma(x, y)$. Furthermore, we consider the weight function w on edges defined by $w(e) = i + 1$ if e is at graph distance i from o. Observe that the length of any of the paths $\gamma(x, y)$ is at most

$$2 \sum_{0}^{2n-1} \frac{1}{i+1} \leq 2 \log(2n+1).$$

Also, the number of times a given edge e at distance i from o is used can be bounded as follows.

$$\begin{aligned} \#\{(x, y) : \gamma(x, y) \ni e\} &\leq (2(n+1)^2 - 1) \times \#\{z : \gamma(z, o) \ni e\} \\ &\leq 4(n+1)^2(2(n+1)^2 - 1)/(i+1). \end{aligned}$$

Hence, The constant A in Theorem 3.2.3 satisfies

$$\begin{aligned} A &\leq \frac{4 \max_{x,y} |\gamma(x, y|_w}{2(n+1)^2 - 1} \max_e \{w(e) \#\{(x, y) : \gamma(x, y) \ni e\}\} \\ &\leq 16(n+1)^2 \log(2n+1). \end{aligned}$$

This yields $\lambda \geq (16(n+1)^2 \log(2n+1))^{-1}$. To see that this is the right order of magnitude, use the test function f defined by $f(x) = \mathrm{sgn}(x)\log(1 + d(0, x))$ where $\mathrm{sgn}(x)$ is $1, 0$ or -1 depending on whether the sum of the coordinates of x is positive 0 or negative. This function has $\pi(f) = 0$,

$$\mathrm{Var}_\pi(f) = \|f\|_2^2 \geq \frac{n(n+1)}{2(n+1)^2 - 1}[\log(n+1)]^2$$

and

$$\begin{aligned}
\mathcal{E}(f,f) &\leq \frac{1}{2[2(n+1)^2-1]} \sum_{i=0}^{2n-1} [(i+1) \wedge (2n-i+1)] |\log(i+2) - \log(i+1)|^2 \\
&\leq \frac{1}{2(n+1)^2-1} \sum_{i=0}^{n-1} \frac{1}{i+1} \\
&\leq \frac{\log(n+1)}{2(n+1)^2-1}.
\end{aligned}$$

Hence, $\lambda \leq [n(n+1)\log(n+1)]^{-1}$. Collecting the results we see that the spectral gap of the n-dog satifies

$$\frac{1}{16(n+1)^2 \log(2n+1)} \leq \lambda \leq \frac{1}{n(n+1)\log(n+1)}.$$

One can convince oneself that there is no choice of paths such that Theorem 3.2.1 give the right order of magnitude. In fact the best that Theorem 3.2.1 gives in this case is $\lambda \geq c/n^3$. The above problem (and its solution) generalizes to any fixed dimension d. For any $d \geq 3$, the corresponding spectral gap satisfies $c_1(d)/n^d \leq \lambda \leq c_2(d)/n^d$.

In Theorems 3.2.1, 3.2.3, exactly one path $\gamma(x, y)$ is used for each pair (x, y). In certain situations it is helpful to allow the use of more than one path from x to y. To this end we introduce the notion of flow.

Definition 3.2.4 *Let (K, π) be an irreducible Markov chain on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set. A flow is non-negative function on the path set Γ,*

$$\phi : \Gamma \to [0, \infty[$$

such that

$$\forall x, y \in \mathcal{X},\ x \neq y, \quad \sum_{\gamma \in \Gamma(x,y)} \phi(\gamma) = \pi(x)\pi(y).$$

Theorem 3.2.5 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set and ϕ be a flow. Then $\lambda \geq 1/A(\phi)$ where*

$$A(\phi) = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{\gamma \in \Gamma: \\ \gamma \ni e}} |\gamma| \phi(\gamma) \right\}.$$

PROOF: This time, for each (x,y) and each $\gamma \in \Gamma(x,y)$ write

$$|f(y)-f(x)|^2 \le |\gamma| \sum_{e\in\gamma} |df(e)|^2.$$

Then

$$|f(y)-f(x)|^2 \pi(x)\pi(y) \le \sum_{\gamma\in\Gamma(x,y)} |\gamma| \sum_{e\in\gamma} |df(e)|^2 \phi(\gamma).$$

Complete the proof as for Theorem 3.2.1.

EXAMPLE 3.2.6: Consider the hypercube $\{0,1\}^n$ with the chain $K(x,y)=0$ unless $|x-y|=1$ in which case $K(x,y)=1/n$. Consider the set $\mathcal{G}(x,y)$ of all geodesic paths from x to y. Define a flow ϕ by setting

$$\phi(\gamma) = \begin{cases} [2^{2n}\#\mathcal{G}(x,y)]^{-1} & \text{if } \gamma\in\mathcal{G}(x,y) \\ 0 & \text{otherwise.} \end{cases}$$

Then $A(\phi)=\max_e A(\phi,e)$ where

$$A(\phi,e) = n2^n \sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}} |\gamma|\phi(\gamma).$$

Using the symmetries of the hypercube, we observe that $A(\phi,e)$ does not depend on e. Summing over the $n2^n$ oriented edges yields

$$\begin{aligned} A(\phi,e) &= \sum_{e\in\mathcal{A}} \sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}} |\gamma|\phi(\gamma) \\ &= \sum_{\gamma} |\gamma|^2\phi(\gamma) \le n^2. \end{aligned}$$

This example generalizes as follows.

Corollary 3.2.6 *Assume that there is a group G which acts on $\mathcal{X}$ and such that*

$$\pi(gx)=\pi(x),\ \ Q(gx,gy)=Q(x,y).$$

Let $\mathcal{A}$ be an adapted edge set such that $(x,y)\in\mathcal{A} \Rightarrow (gx,gy)\in\mathcal{A}$. Let $\mathcal{A}=\bigcup_1^k \mathcal{A}_i$, be the partition of $\mathcal{A}$ into transitive classes for this action. Then $\lambda \ge 1/A$ where

$$A = \max_{1\le i\le k} \left\{ \frac{1}{|\mathcal{A}_i| Q_i} \sum_{x,y} d(x,y)^2 \pi(x)\pi(y) \right\}.$$

Here $|\mathcal{A}_i| = \#\mathcal{A}_i$, $Q_i = Q(e_i)$ with $e_i\in\mathcal{A}_i$, and $d(x,y)$ is the graph distance between x and y.

PROOF: Consider the set $\mathcal{G}(x,y)$ of all geodesic paths from x to y. Define a flow ϕ by setting

$$\phi(\gamma)=\begin{cases}\pi(x)\pi(y)/\#\mathcal{G}(x,y) & \text{if } \gamma\in\mathcal{G}(x,y)\\ 0 & \text{otherwise.}\end{cases}$$

Then $A(\phi)=\max_e A(\phi,e)$ where

$$A(\phi,e)=\frac{1}{Q(e)}\sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}}|\gamma|\phi(\gamma).$$

By hypothesis, $A(\phi,e_i)=A_i(\phi)$ does not depend on $e_i\in\mathcal{A}_i$. Indeed, if $g\gamma$ denote the image of the path γ under the action of $g\in G$, we have $|g\gamma|=|\gamma|$, $\phi(g\gamma)=\phi(\gamma)$. Summing for each $i=1,\dots,k$ over all the oriented edges in $\mathcal{A}_i$, we obtain

$$\begin{aligned}A(\phi,e_i) &= \frac{1}{|\mathcal{A}_i|Q_i}\sum_{e\in\mathcal{A}_i}\sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}}|\gamma|\phi(\gamma)\\ &= \frac{1}{|\mathcal{A}_i|Q_i}\sum_{e\in\mathcal{A}_i}\sum_{x,y}\sum_{\substack{\gamma\in\mathcal{G}(x,y):\\ \gamma\ni e}}\frac{d(x,y)\pi(x)\pi(y)}{\#\mathcal{G}(x,y)}\\ &\le \frac{1}{|\mathcal{A}_i|Q_i}\sum_{x,y}N_i(x,y)d(x,y)\pi(x)\pi(y)\end{aligned}$$

where

$$N_i(x,y)=\max_{\gamma\in\mathcal{G}(x,y)}\#\{e\in\mathcal{A}_i:\gamma\ni e\}.$$

That is, $N_i(x,y)$ is the maximal number of edges of type i used in a geodesic path from x to y. In particular, $N_i(x,y)\le d(x,y)$ and the announced result follows.

EXAMPLE 3.2.7: Let $\mathcal{X}$ be the set of all k-subsets of a set with n elements. Assume $k\le n/2$. Consider the graph with vertex set $\mathcal{X}$ and an edge from x to y if $\#(x\cap y)=k-2$. This is a regular graph with degree $k(n-k)$. The simple random walk on this graph has kernel

$$K(x,y)=\begin{cases}1/[k(n-k)] & \text{if } \#(x\cap y)=k-2\\ 0 & \text{otherwise}\end{cases}$$

and stationary measure $\pi\equiv\binom{n}{k}^{-1}$. It is clear that the symmetric group S_n acts transitively on the edge set of this graph and preserves K and π. Here there is only one class of edges, $|\mathcal{A}|=\binom{n}{k}n(n-k)$, $Q=|\mathcal{A}|^{-1}$. Therefore Corollary 3.2.6 yields $\lambda\ge 1/A$ with

$$A = \frac{1}{|\mathcal{A}|Q}\sum_{x,y}d(x,y)^2\pi(x)\pi(y)$$

$$\begin{aligned}
&= \frac{1}{\binom{n}{k}^2}\sum_1^k \ell^2 \binom{n}{k}\binom{k}{\ell}\binom{n-k}{\ell} \\
&= \frac{1}{\binom{n}{k}}\sum_1^k \ell^2 \binom{k}{\ell}\binom{n-k}{\ell} \\
&= \frac{k(n-k)}{\binom{n}{k}}\sum_1^k \binom{k-1}{\ell-1}\binom{n-k-1}{\ell-1} \\
&= \frac{k(n-k)}{\binom{n}{k}}\binom{n-2}{k-1} = \frac{k^2(n-k)^2}{n(n-1)}.
\end{aligned}$$

Hence

$$\lambda \geq \frac{n(n-1)}{k^2(n-k)^2}.$$

Here we have used the fact that the number of pair (x,y) with $d(x,y)=\ell$ is $\binom{n}{k}\binom{k}{\ell}\binom{n-k}{\ell}$ to obtain the second equality. Also, the equality

$$\sum_1^k \binom{k-1}{\ell-1}\binom{n-k-1}{\ell-1} = \binom{n-2}{k-1}$$

can be proved by counting in two different ways how to draw $k-1$ marked balls from an urn containing $n-2$ balls, $k-1$ of them being red, the others black. Indeed, $\binom{k-1}{\ell-1}\binom{n-k-1}{\ell-1} = \binom{k-1}{k-\ell}\binom{n-k-1}{\ell-1}$ is the number of different ways to pick $k-1$ balls, $\ell-1$ of which are black. The true value of λ is $n/[k(n-k)]$. See [34].

EXAMPLE 3.2.8: Let $\mathcal{X}$ be the set of all n-subsets of $\{0,\dots,2n-1\}$. Consider the graph with vertex set $\mathcal{X}$ and an edge from x to y if $\#(x\cap y)=n-2$ and $0\in x\oplus y$ where $x\oplus y = x\cup y\setminus x\cap y$ is the symmetric difference of x and y. This is a regular graph with degree n. The simple random walk on this graph has kernel

$$K(x,y)=\begin{cases} 1/n & \text{if } \#(x\cap y)=n-2 \text{ and } 0\in x\oplus y\\ 0 & \text{otherwise}\end{cases}$$

and stationary measure $\pi \equiv \binom{2n}{n}^{-1}$. This process can be described informally as follows: Let x be subset of $\{0,\dots,2n-1\}$ having n elements. If $0\in x$, pick an element a uniformly at random in the complement of x and move to $y=(x\setminus\{0\})\cup\{a\}$, that is, replace 0 by a. If $0\notin x$, pick an element a uniformly at random in x and move to $y=(x\setminus\{a\})\cup\{0\}$, that is, replace a by 0.

It is clear that the symmetric group S_{2n-1} which fixes 0 and acts on $\{1,\dots,2n-1\}$ also acts on this graph and preserves K and π. This action is not transitive on edges. There are two transitive classes $\mathcal{A}_1,\mathcal{A}_2$ of edges depending on whether, for an edge (x,y), $0\in x$ or $0\in y$. Clearly

$$|\mathcal{A}_1|=|\mathcal{A}_2|=\binom{2n}{n}n,\;\; Q_1=Q_2=|\mathcal{A}|^{-1}=(2|\mathcal{A}_1|)^{-1}.$$

If x and y differ by exactly ℓ elements, the distance between x and y is 2ℓ if $0 \notin x \oplus y$ and $2\ell - 1$ if $0 \in x \oplus y$. Using this and a computation similar to the one in Example 3.2.7, we see that the constant A in Corollary 3.2.6 is bounded by

$$\begin{aligned} A &= \frac{1}{|\mathcal{A}_1|Q_1} \sum_{x,y} d(x,y)^2 \pi(x)\pi(y) \\ &\leq \frac{8}{\binom{2n}{n}^2} \sum_1^n \ell^2 \binom{2n}{n}\binom{n}{\ell}^2 \\ &= \frac{4n^2}{2n-1}. \end{aligned}$$

Hence $\lambda \geq (2n-1)/(4n^2)$. This can be slightly improved if we use the $N_i(x,y)$'s introduced in the proof of Corollary 3.2.6. Indeed, this proof shows that $\lambda \geq 1/A'$ with

$$A' = \max_i \left\{ \frac{1}{|\mathcal{A}_i|Q_i} \sum_{x,y} N_i(x,y) d(x,y) \pi(x)\pi(y) \right\}$$

where $N_i(x,y)$ is the maximal number of edges of type i used in any geodesic path from x to y. In the present case, if $x \oplus y = \ell$, then the distance between x and y is atmost 2ℓ with atmost ℓ edges of each of the two types. Hence, $A' \leq 2n^2/(2n-1)$ and $\lambda \geq (2n-1)/(2n^2)$. This bound is of the correct order of magnitude is $1/n$. See the end of Section 4.2.

Corollary 3.2.7 *Assume that $\mathcal{X} = G$ is a finite group with generating set $S = \{g_1, \ldots, g_s\}$. Set $K(x,y) = |S|^{-1}\mathbf{1}_S(x^{-1}y)$, $\pi \equiv 1/|G|$. Then*

$$\lambda(K) \geq \frac{1}{2|S|D^2}$$

where D is the diameter of the Cayley graph $(G, S \cup S^{-1})$. If S is symmetric, i.e., $S = S^{-1}$, then

$$\lambda(K) \geq \frac{1}{|S|D^2}.$$

PROOF: The action of the group G on its itself by left translation preserves K and π. Hence it also preserves Q. We set

$$\mathcal{A} = \left\{(x, xs) : x \in G, s \in S \cup S^{-1}\right\}.$$

There are at most $s = 2|S|$ classes of oriented edges (corresponding to the distinct elements of $S \cup S^{-1}$) and each class contains at least $|G|$ distinct edges. If S is symmetric (that is $g \in S \Rightarrow g^{-1} \in S$) then $1/Q(e) = |S||G|$ whereas if S is not symmetric, $|S||G| \leq 1/Q(e) \leq 2|S||G|$. The results now follow from Corollary 3.2.6. Slightly better bounds are derived in [24].

Corollary 3.2.8 *Assume that $\mathcal{X} = G$ is a finite group with generating set $S = \{g_1, \ldots, g_s\}$. Set $K(x,y) = |S|^{-1}\mathbf{1}_S(x^{-1}y)$, $\pi \equiv 1/|G|$. Assume that there is a subgroup H of the group of automorphisms of G which preserves S and acts transitively on S. Then*

$$\lambda(K) \geq \frac{1}{2D^2}$$

where D is the diameter of the Cayley graph $(G, S \cup S^{-1})$. If S is symmetric, i.e., $S = S^{-1}$, or if H acts transitively on $S \cup S^{-1}$, then

$$\lambda(K) \geq \frac{1}{D^2}.$$

These results apply in particular when S is a conjugacy class.

Proof: Let $e_i = (x_i, x_i s_i) \in \mathcal{A}$, $x_i \in G$, $s_i \in S \cup S^{-1}$, $i = 1, 2$ be two edges. If $s_1, s_2 \in S$, there exists $\sigma \in H$ such that $\sigma(s_1) = s_2$. Set $\sigma(x_1) = y_1$. Then $z \to x_2 y_1^{-1}\sigma(z)$ is an automorphism of G which send x_1 to x_2 and $x_1 s_1$ to $x_2 s_2$. A similar reasoning applies if $s_1, s_2 \in S^{-1}$. Hence there are atmost two transitive classes of edges. If there are two classes, $(x, xs) \to (x, xs^{-1})$ establishes a bijection between them. Hence $|\mathcal{A}_1| = |\mathcal{A}_2| = |\mathcal{A}|/2$. Hence the desired results follow from Corollary 3.2.6.

Example 3.2.9: Let $\mathcal{X} = S_n$ be the symmetric group on n objects. Let $K(x,y) = 0$ unless $y = x\sigma_i$ with $\sigma_i = (1, i)$ and $i = \{2, \ldots, n\}$, in which case $K(x,y) = 1/(n-1)$. Decomposing any permutation θ in to disjoint cycles shows that θ is a product of at most n transpositions. Further more, any transposition (i,j) can be written as $(i,j) = (1,i)(1,j)(1,i)$. Hence any permutation is a product of at most $3n$ σ_i's and Corollary 3.2.7 yields $\lambda \geq 9n^3$. However, the subgroup $S_{n-1}(1) \subset S_n$ of the permutations that fixe 1 acts by conjugaison on S_n. Set $\psi_h : x \to hxh^{-1}$, $h \in S_{n-1}(1)$ and $H = \{\psi_h : S_n \to S_n : \ h \in S_{n-1}(1)\}$. This group of automorphisms of S_n acts transitively on $S = \{\sigma_i : i \in \{2, \ldots, n\}\}$. Indeed, for $2 \leq i, j \leq n$, $h = (i,j) \in S_{n-1}(1)$ satisfies $\psi_h(\sigma_i) = \sigma_j$. Hence Corollary 3.2.8 gives the improved bound $\lambda \geq 9n^2$. The right answer is that $\lambda = 1/n$ by Fourier analysis [42].

To conclude this section we observe that there is no reason why we should choose between using a weight function as in Theorem 3.2.3 or using a flow as in Theorem 3.2.5. Furthermore we can consider more general weight functions

$$w : \Gamma \times \mathcal{A} \to (0, \infty)$$

where the weight $w(\gamma, e)$ of an edge also depends on which path γ we are considering. Again, we set $|\gamma|_w = \sum_{e \in \gamma} w(\gamma, e)^{-1}$. Then we have

Theorem 3.2.9 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set, w a generalized weight function and*

ϕ a flow. Then $\lambda \geq 1/A(w,\phi)$ where

$$A(w,\phi) = \max_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}} w(\gamma,e)|\gamma|_w \phi(\gamma) \right\}.$$

3.3 Isoperimetry

3.3.1 Isoperimetry and spectral gap

It is well known that spectral gap bounds can be obtained through isoperimetric inequalities via the so-called Cheeger's inequality introduced in a different context in Cheeger [12]. See Alon [5], Alon and Milman [6], Sinclair [71, 72], Diaconis and Stroock [35], Kannan [56], and the earlier references given there. See also [58]. This section presents this technique. It emphasizes the fact that isoperimetric inequalities are simply ℓ^1 version of Poincaré inequalities. It follows that in most circumstances it is possible and preferable to work directly with Poincaré inequalities if the ultimate goal is to bound the spectral gap. Diaconis and Stroock [35] compare bounds using Theorems 3.2.1, 3.2.3, and bounds using Cheeger's inequality. They find that, most of the time, bounds using Cheeger's inequality can be tightned by appealing directly to a Poincaré inequality.

Definition 3.3.1 *The "boundary" ∂A of a set $A \subset \mathcal{X}$ is the set*

$$\partial A = \{e = (x,y) \in \mathcal{X} \times \mathcal{X} :\ x \in A, y \in A^c \text{ or } x \in A^c, y \in A\}.$$

Thus, the boundary is the set of all pairs connecting A and A^c.

Given a Markov chain (K,π), the measure of the boundary ∂A of $A \subset X$ is

$$\begin{aligned} Q(\partial A) &= \frac{1}{2} \sum_{(x,y)\in\partial A} (K(x,y)\pi(x) + K(y,x)\pi(y)) \\ &= \sum_{x\in A, y\in A^c} (K(x,y)\pi(x) + K(y,x)\pi(y)). \end{aligned}$$

The "boundary" ∂A is a rather large boundary and does not depend on the chain (K,π) under consideration. However, only the portion of ∂A that has positive Q-measure will be of interest to us so that we could as well have required that the edges in ∂A satisfy $Q(e) > 0$.

Definition 3.3.2 *The isoperimetric constant of the chain (K,π) is defined by*

$$I = I(K,\pi) = \min_{\substack{A\subset\mathcal{X}:\\ \pi(A)\leq 1/2}} \left\{ \frac{Q(\partial A)}{\pi(A)} \right\}. \tag{3.3.1}$$

Let us specialize this definition to the case where (K,π) is the simple random walk on an r-regular graph $(\mathcal{X},\mathcal{A})$. Then, $K(x,y) = 1/r$ if x,y are neighbors and

$\pi(x) \equiv 1/|\mathcal{X}|$. Hence $Q(e) = 1/(r|\mathcal{X}|)$ if $e \in \mathcal{A}$. Define the geometric boundary of a set A to be

$$\partial_* A = \{(x, y) \in \mathcal{A} : x \in A, y \in A^c\}.$$

Then

$$I = \min_{\substack{A \subset \mathcal{X}: \\ \pi(A) \leq 1/2}} \left\{ \frac{Q(\partial A)}{\pi(A)} \right\} = \frac{2}{r} \min_{\substack{A \subset \mathcal{X}: \\ \# A \leq \# \mathcal{X}/2}} \left\{ \frac{\# \partial_* A}{\# A} \right\}.$$

Lemma 3.3.3 *The constant I satisfies*

$$I = \min_f \left\{ \frac{\sum_e |df(e)| Q(e)}{\min_\alpha \sum_x |f(x) - \alpha| \pi(x)} \right\}.$$

Here the minimum is over all non-constant fonctions f.

It is well known and not too hard to prove that

$$\min_\alpha \sum_x |f(x) - \alpha| \pi(x) = \sum_x |f(x) - \alpha_0| \pi(x)$$

if and only if α_0 satisfies

$$\pi(f > \alpha_0) \leq 1/2 \text{ and } \pi(f < \alpha_0) \leq 1/2$$

i.e., if and only if α_0 is a median.

Proof: Let J be the right-hand side in the equality above. To prove that $I \geq J$ it is enough to take $f = 1_A$ in the definition of J. Indeed,

$$\sum_e |d1_A(e)| Q(e) = Q(\partial A), \quad \sum_x 1_A(x) \pi(x) = \pi(A).$$

We turn to the proof of $J \geq I$. For any non-negative function f, set $F_t = \{f \geq t\}$ and $f_t = 1_{F_t}$. Then observe that $f(x) = \int_0^\infty f_t(x) dt$,

$$\pi(f) = \int_0^\infty \pi(F_t) dt$$

and

$$\sum_e |df(e)| Q(e) = \int_0^\infty Q(\partial F_t) dt. \tag{3.3.2}$$

This is a discrete version of the so-called co-area formula of geometric measure theory. The proof is simple. Write

$$\begin{aligned} \sum_e |df(e)| Q(e) &= 2 \sum_{\substack{e=(x,y) \\ f(y) > f(x)}} (f(y) - f(x)) Q(e) \\ &= 2 \sum_{\substack{e=(x,y) \\ f(y) > f(x)}} \int_{f(x)}^{f(y)} Q(e) dt \\ &= 2 \int_0^\infty \sum_{\substack{e=(x,y) \\ f(y) \geq t > f(x)}} Q(e) dt \\ &= \int_0^\infty Q(\partial F_t) dt. \end{aligned}$$

Given a function f, let α be such that $\pi(f > \alpha) \leq 1/2$, $\pi(f < \alpha) \leq 1/2$ and set $f_+ = (f - \alpha) \vee 0$, $f_- = -[(f - \alpha) \wedge 0]$. Then, $f_+ + f_- = |f - \alpha|$ and $|df(e)| = |df_+(e)| + |df_-(e)|$. Setting $F_{\pm,t} = \{x : f_\pm(x) \geq t\}$, using (3.3.2) and the definition of I, we get

$$\begin{aligned}
\sum_e |df(e)|Q(e) &= \sum_e |df_+(e)|Q(e) + \sum_e |df_-(e)|Q(e) \\
&= \int_0^\infty Q(\partial F_{+,t})dt + \int_0^\infty Q(\partial F_{-,t})dt \\
&\geq I \int_0^\infty (\pi(F_{+,t}) + \pi(F_{-,t}))dt \\
&= I \sum_x (f_+(x) + f_-(x))\pi(x) \\
&= I \sum_x |f(x) - \alpha|\pi(x).
\end{aligned}$$

This proves that $J \geq I$.

There is an alternative notion of isoperimetric constant that is sometimes used in the literature.

Definition 3.3.4 *Define the isoperimetric constant I' of the chain (K, π) by*

$$I' = I'(K, \pi) = \min_{A \subset \mathcal{X}} \left\{ \frac{Q(\partial A)}{2\pi(A)(1 - \pi(A))} \right\}. \tag{3.3.3}$$

Observe that $I/2 \leq I' \leq I$.

Lemma 3.3.5 *The constant I' is also given by*

$$I' = \min_f \left\{ \frac{\sum_e |df(e)|Q(e)}{\sum_x |f(x) - \pi(f)|\pi(x)} \right\}$$

where the minimum is taken over all non-constant functions f.

PROOF: Setting $f = 1_A$ in the ratio appearing above shows that the left-hand side is not smaller than the right-hand side. To prove the converse, set $f_+ = f \vee 0$, and $F_t = \{x : f_+(x) \geq t\}$. As in the proof of Lemma 3.3.3, we obtain

$$\sum_e |df_+(e)|Q(e) \geq 2I' \int_0^\infty \pi(F_t)(1 - \pi(F_t))dt.$$

Now,

$$\begin{aligned}
2\pi(F_t)(1 - \pi(F_t)) &= \sum_x |1_{F_t}(x) - \pi(1_{F_t})|\pi(x) \\
&= \max_{\substack{g;\pi(g)=0 \\ \min_\alpha |g-\alpha| \leq 1}} \sum_x 1_{F_t}(x)\, g(x)\pi(x).
\end{aligned}$$

Here, we have used the fact that, for any function u,

$$\sum_x |u(x) - \pi(u)|\pi(x) = \max_{\substack{g;\pi(g)=0 \\ \min_\alpha |g-\alpha|\le 1}} \sum_x u(x)\, g(x)\pi(x).$$

See [68]. Thus, for any g satifying $\pi(g) = 0$ and $\min_\alpha |g - \alpha| \le 1$,

$$\begin{aligned}\sum_e |df_+(e)|Q(e) &\geq I' \sum_x \left(\int_0^\infty 1_{F_t}(x)\, dt\right) g(x)\pi(x) \\ &\geq I' \sum_x f_+(x)g(x)\pi(x).\end{aligned}$$

The same reasoning applies to $f_- = -[f \wedge 0]$ so that, for all g as above,

$$\sum_e |df_-(e)|Q(e) \geq I' \sum_x f_-(x)g(x)\pi(x).$$

Adding the two inequalities, and taking the supremum over all allowable g, we get

$$\sum_e |df(e)|Q(e) \geq I' \sum_x |f(x) - \pi(f)|\pi(x)$$

which is the desired inequality.

Lemmas 3.3.3 and 3.3.5 shows that the argument used in the proof of Theorem 3.2.1 can be used to bound I and I' from below.

Theorem 3.3.6 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set. For each $(x, y) \in \mathcal{X} \times \mathcal{X}$ choose exactly one path $\gamma(x, y)$ in $\Gamma(x, y)$. Then $I \geq I' \geq 1/B$ where*

$$B = \max_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y\in\mathcal{X}: \\ \gamma(x,y)\ni e}} \pi(x)\pi(y) \right\}.$$

PROOF: For each $(x, y) \in \mathcal{X} \times \mathcal{X}$, write $f(y) - f(x) = \sum_{e\in\gamma(x,y)} df(e)$ and

$$|f(y) - f(x)| \leq \sum_{e\in\gamma(x,y)} |df(e)|.$$

Multiply by $\pi(x)\pi(y)$ and sum over all x, y to obtain

$$\sum_{x,y} |f(y) - f(x)|\pi(x)\pi(y) \leq \sum_{x,y} \sum_{e\in\gamma(x,y)} |df(e)|\pi(x)\pi(y).$$

This yields

$$\sum_x |f(x) - \pi(f)|\pi(x) \leq B \sum_e |df(e)|Q(e)$$

which implies the desired conclusion. There is also a version of this result using flows as in Theorem 3.2.5.

Lemma 3.3.7 (Cheeger's inequality) *The spectral gap* λ *and the isoperimetric constant* I, I' *defined at* (3.3.1), (3.3.3) *are related by*

$$\frac{I'^2}{8} \leq \frac{I^2}{8} \leq \lambda \leq I' \leq I.$$

Compare with [35], Section 3.C. There, it is proved by a slightly different argument that $h^2/2 \leq \lambda < 2h$ where $h = I/2$. This is the same as $I^2/8 \leq \lambda \leq I$.

Proof: For the upper bound use the test functions $f = 1_A$ in the definition of λ. For the lower bound, apply

$$\sum_e |df(e)| Q(e) \geq I \min_\alpha \sum_x |f(x) - \alpha| \pi(x)$$

to the function $f = |g-c|^2 \mathrm{sgn}(g-c)$ where g is an arbitrary function and $c = c(g)$ is a median of g so that $\sum_x |f(x) - \alpha|\pi(x)$ is minimum for $\alpha = 0$. Then, for $e = (x, y)$,

$$|df(e)| \leq |dg(e)|(|g(x) - c| + |g(y) - c|)$$

because $|a^2 - b^2| = |a - b|(|a| + |b|)$ if $ab \geq 0$ and $a^2 + b^2 \leq |a - b|(|a| + |b|)$ if $ab < 0$. Hence

$$\begin{aligned}
\sum_e |df(e)| Q(e) &\leq \sum_{e=(x,y)} |dg(e)|(|g(x) - c| + |g(y) - c|) Q(e) \\
&\leq \left(\sum_e |dg(e)|^2 Q(e) \right)^{1/2} \times \\
&\quad \left(2 \sum_{x,y} (|g(x) - c|^2 + |g(y) - c|^2) \pi(x) K(x, y) \right)^{1/2} \\
&= (8\mathcal{E}(g, g))^{1/2} \left(\sum_x |g(x) - c|^2 \pi(x) \right)^{1/2}.
\end{aligned}$$

Hence

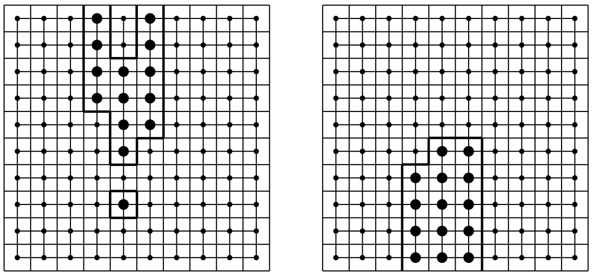
$$\begin{aligned}
I \sum_x |g(x) - c|^2 \pi(x) &= I \min_\alpha \sum_x |f(x) - \alpha|^2 \pi(x) \\
&\leq \sum_e |df(e)| Q(e) \\
&\leq (8\mathcal{E}(g, g))^{1/2} \left(\sum_x |g(x) - c|^2 \pi(x) \right)^{1/2}.
\end{aligned}$$

and

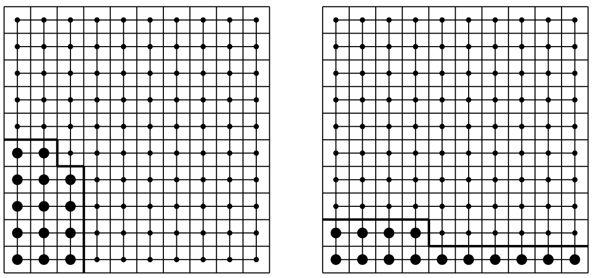
$$I^2 \mathrm{Var}_\pi(g) \leq I^2 \sum_x |g(x) - c|^2 \pi(x) \leq 8\mathcal{E}(g, g).$$

for all functions g. This proves the desired lower bound.

EXAMPLE 3.3.1: Let $\mathcal{X} = \{0, \dots, n\}^2$ be the vertex set of a square grid of side n. Hence, the edge set $\mathcal{A}$ is given by $\mathcal{A} = \{(x, y) \in \mathcal{X}^2 : |x-y| = 1\}$ where $|x-y|$ denote either the Euclidian distance or simply $\sum_i |x_i - y_i|$ (it does not matter which). Define $K(x, y)$ to be zero if $|x - y| \geq 1$, $K(x, y) = 1/4$ if $|x - y| = 1$, and $K(x, x) = 0$, $1/4$ or $1/2$ depending on whether x is interior, on a side, or a corner of $\mathcal{X}$. The uniform distribution $\pi \equiv 1/(n + 1)^2$ is the reversible measure of K. To have a more geometric interpretation of the boundary, we view each vertex in $\mathcal{X}$ as the center of a unit square as in the figure below.



Then, for any subset $A \subset \mathcal{X}$, $\pi(A)$ is proportional to the surface of those unit squares with center in A. Call $\mathbf{A}$ the union of those squares (viewed as a subset of the plane). Now $Q(\partial A)$ is proportional to the length of the interior part of the boundary of $\mathbf{A}$. It is not hard to see that pushing all squares in each column down to the bottom leads to a set $\mathbf{A}^{\downarrow}$ with the same area and smaller boundary.



Similarly, we can push things left. Then consider the upper left most unit square. It is easy to see that moving it down to the left bottom most free space does not increase the boundary. Repeating this operation as many times as possible shows that, given a number N of unit squares, the smallest boundary is obtained for the set formed with $[N/(n+1)]$ bottom raws and the $N - (n+1)[N/(n+1)]$ left most squares of the $([N/(n + 1)] + 1)^{th}$ raw. Hence, we have

$$\frac{Q(\partial A)}{\pi(A)} = \begin{cases} \frac{N+1}{4N} & \text{if } \#A = N \leq n+1 \\ \frac{n+2}{4N} & \text{if } n+1 \leq \#A = N \text{ and } \#A \text{ does not divide } n+1 \\ \frac{n+1}{4N} & \text{if } \#A = N = k(n+1). \end{cases}$$

Theorem 3.3.8 *For the natural walk on the square grid $\mathcal{X} = \{0, \ldots, n\}^2$ the isoperimetric constants I, I' are given by*

$$I = \begin{cases} \frac{1}{2(n+1)} & \text{if } n+1 \text{ is even} \\ \frac{1}{2n} & \text{if } n+1 \text{ is odd.} \end{cases} \quad I' = \begin{cases} \frac{1}{2(n+1)} & \text{if } n+1 \text{ is even} \\ \frac{1}{2n(1+(n+1)^{-2})} & \text{if } n+1 \text{ is odd.} \end{cases}$$

Using Cheeger's inequality yields

$$\lambda \geq \frac{1}{32(n+1)^2}.$$

This is of the right order of magnitude.

Example 3.3.2: For comparison, consider the example of the "n-dog". That is, two square grids as above with one corner o identified. In this case, it is clear that the ratio $Q(\partial A)/\pi(A)$ (with $\pi(A) \leq 1/2$) is smallest for A one of the two squares minus o. Hence

$$I(n\text{-dog}) = \frac{1}{2[(n+1)^2 - 1]}.$$

In this case Cheeger's inequality yields

$$\lambda(n\text{-dog}) \geq \frac{1}{32(n+1)^4}.$$

This is far off from the right order of magnitude $1/(n^2 \log n)$ which was found using Theorem 3.2.3.

The proof of Theorem 3.3.8 works as well in higher dimension and for rectangular boxes.

Theorem 3.3.9 *For the natural walk on the parallelepiped*

$$\mathcal{X} = \{0, \ldots, n_1\} \times \ldots \times \{0, \ldots, n_d\}$$

with $n_1 = \max n_i$, the isoperimetric constants I, I' satisfy

$$I \geq I' \geq \frac{1}{d(n_1 + 1)}.$$

In this case, Cheeger's inequality yields a bound which is off by a factor of $1/d$.

The above examples must not lead the reader to believe that, generaly speaking, isoperimetric inequalities are easy to prove or at least easier to prove than Poincaré inequalities. It is the case in some examples as the ones above whose geometry is really simple. There are other examples where the spectral gap is known exactly (e.g., by using Fourier analysis) but where even the order of magnitude of the isoperimetric constant I is not known. One such example is provided by the walk on the symmetric group S_n with $K(x, y) = 2/n(n-1)$ if x and y differ by a transposition and $K(x, y) = 0$ otherwise. For this walk $\lambda = 2/(n-1)$ and, by Cheeger's inequality, $2/(n-1) \leq I \leq 4/(n-1)^{1/2}$.

3.3.2 Isoperimetry and Nash inequalities

The goal of this section is to prove the following result.

Theorem 3.3.10 *Assume that* (K, π) *satisfies*

$$\pi(A)^{(d-1)/d} \leq S\left(Q(\partial A) + \frac{1}{R}\pi(A)\right) \tag{3.3.4}$$

for all $A \subset \mathcal{X}$ *and some constants* $d \geq 1$, $S, R > 0$. *Then*

$$\forall\, g,\ \ \|g\|_{d/(d-1)} \leq S\left(\sum_e |dg(e)|Q(e) + \frac{1}{R}\|g\|_1\right) \tag{3.3.5}$$

and

$$\forall\, g,\ \ \|g\|_2^{2(1+2/d)} \leq 16\, S^2\left(\mathcal{E}(g,g) + \frac{1}{8R^2}\|g\|_2^2\right)\|g\|_1^{4/d}. \tag{3.3.6}$$

Proof: Since $|d|g|(e)| \leq |dg(e)|$ it suffices to prove the result for $g \geq 0$. Write $g = \int_0^\infty g_t dt$ where $g_t = 1_{G_t}$, $G_t = \{g \geq t\}$, and set $q = d/(d-1)$. Then

$$\begin{aligned}
\|g\|_q &\leq \int_0^\infty \|g_t\|_q dt = \int_0^\infty \pi(G_t)^{1/q} dt \\
&\leq S\int_0^\infty \left(Q(\partial G_t) + \frac{1}{R}\pi(G_t)\right) dt \\
&= S\left(\sum_e |dg(e)|Q(e) + \frac{1}{R}\|g\|_1\right).
\end{aligned}$$

The first inequality uses Minkowski's inequality. The second inequality uses (3.3.4). The last inequality uses the co-area formula (3.3.2). This proves (3.3.5). It is easy to see that (3.3.5) is in fact equivalent to (3.3.4) (take $g = 1_A$).

To prove (3.3.6), we observe that

$$\sum_e |dg^2(e)|Q(e) \leq [8\mathcal{E}(g,g)]^{1/2}\|g\|_2.$$

Indeed,

$$\begin{aligned}
\sum_e |dg^2(e)|Q(e) &= \sum_{e=(x,y)} |dg(e)||g(x)+g(y)|Q(e) \\
&\leq \left(\sum_e |dg(e)|^2 Q(e)\right)^{1/2} \times \\
&\qquad \left(2\sum_{x,y}(|g(x)|^2 + |g(y)|^2)\pi(x)K(x,y)\right)^{1/2} \\
&= (8\mathcal{E}(g,g))^{1/2}\left(\sum_x |g(x)|^2\pi(x)\right)^{1/2}.
\end{aligned}$$

Thus, (3.3.5) applied to g^2 yields

$$\|g\|_{2q}^2 \le S\left([8\mathcal{E}(g,g)]^{1/2}\|g\|_2 + \frac{1}{R}\|g\|_2^2\right)$$

with $q = d/(d-1)$. The Hölder inequality

$$\|g\|_2 \le \|g\|_1^{1/(1+d)}\|g\|_{2q}^{d/(1+d)}$$

and the last inequality let us bound $\|g\|_2$ by

$$\left(S\left([8\mathcal{E}(g,g)]^{1/2}\|g\|_2 + \frac{1}{R}\|g\|_2^2\right)\right)^{1/[2(1+d)]} \|g\|_1^{1/(1+d)}$$

We raise this to the power $2(1+d)/d$ and divide by $\|g\|_2$ to get

$$\|g\|_2^{(1+2/d)} \le S\left([8\mathcal{E}(g,g)]^{1/2} + \frac{1}{R}\|g\|_2\right)\|g\|_1^{2/d}.$$

This yields the desired result.

There is a companion result related to Theorem 2.3.1 and Nash inequalities of type (2.3.1) versus (2.3.3).

Theorem 3.3.11 *Assume that (K,π) satisfies*

$$\pi(A)^{(d-1)/d} \le SQ(\partial A) \tag{3.3.7}$$

for all $A \subset \mathcal{X}$ such that $\pi(A) \le 1/2$. Then

$$\forall\, g \in \ell^2(\pi),\quad \mathrm{Var}_\pi(g)^{(1+2/d)} \le 8S^2\mathcal{E}(g,g)\|f\|_1^{4/d}.$$

Before proving this theorem, let us introduce the isoperimetric constant associated with inequality (3.3.7).

Definition 3.3.12 *The d-dimensional isoperimetric constant of a finite chain (K,π) is defined by*

$$I_d = I_d(K,\pi) = \min_{\substack{A\subset\mathcal{X}:\\ \pi(A)\le 1/2}} \frac{Q(\partial A)}{\pi(A)^{1/q}}$$

where $q = d/(d-1)$.

Observe that $I \ge I_d$ with I the isoperimetric constant defined at (3.3.1) (in fact $I \ge 2^{1/d} I_d$). It may be helpful to specialize this definition to the case where (K,π) is the simple random walk on a r-regular connected symmetric graph $(\mathcal{X},\mathcal{A})$. Then $Q(e) = 1/|\mathcal{A}| = 1/(r|\mathcal{X}|)$, $\pi \equiv 1/|\mathcal{X}|$ and

$$I_d = \frac{2}{r|\mathcal{X}|^{1/d}} \min_{\substack{A\subset\mathcal{X}:\\ \#A\le\#\mathcal{X}/2}} \frac{\#\partial_* A}{[\#A]^{1/q}}$$

where $\partial_* A = \{(x,y) \in \mathcal{A} : x \in A,\ y \notin A\}$.

Lemma 3.3.13 *The isoperimetric constant $I_d(K,\pi)$ is also given by*

$$I_d(K,\pi) = \inf\left\{ \frac{\sum_e |df(e)|Q(e)}{\|f-c(f)\|_q} : f \textit{ non-constant} \right\}$$

where $q = d/(d-1)$ and $c(f)$ denote the smallest median of f.

PROOF: For $f = \mathbf{1}_A$ with $\pi(A) \le 1/2$, $c(f) = 0$ is the smallest median of f. Hence

$$\frac{\sum_e |df(e)|Q(e)}{\|f-c(f)\|_q} = \frac{Q(\partial A)}{\pi(A)^{1/q}}.$$

It follows that

$$\min_f \left\{ \frac{\sum_e |df(e)|Q(e)}{\|f-c(f)\|_q} \right\} \le I_d(K,\pi).$$

To prove the converse, fix a function f and let c be such that $\pi(f > c) \le 1/2$, $\pi(f < c) \le 1/2$. Set $f_+ = (f-c)\vee 0$, $f_- = -[(f-c)\wedge 0]$. Then $f_+ + f_- = |f-c|$ and $|df(e)| = |df_+(e)| + |df_-(e)|$. Setting $F_{\pm,t} = \{x : f_\pm(x) \ge t\}$ and using (3.3.2) we obtain

$$\begin{aligned}
\sum_e |df(e)|Q(e) &\ge \sum_e |df_+(e)|Q(e) + \sum_e |df_-(e)|Q(e) \\
&= \int_0^\infty Q(\partial F_{+,t})dt + \int_0^\infty Q(\partial F_{-,t})dt \\
&\ge I_d \int_0^\infty \left(\pi(F_{+,t})^{1/q} + \pi(F_{-,t})^{1/q} \right) dt.
\end{aligned}$$

Now

$$\pi(F_{\pm,t})^{1/q} = \|\mathbf{1}_{F_{\pm,t}}\|_q = \max_{\|g\|_r \le 1} \langle \mathbf{1}_{F_{\pm,t}}, g\rangle$$

where $1/r + 1/q = 1$. Hence, for any g such that $\|g\|_r \le 1$,

$$\begin{aligned}
\sum_e |df(e)|Q(e) &\ge I_d \int_0^\infty \left(\langle \mathbf{1}_{F_{+,t}}, g\rangle + \langle \mathbf{1}_{F_{-,t}}, g\rangle \right) \\
&= I_d \left(\langle f_+, g\rangle + \langle f_-, g\rangle \right) \\
&= I_d \,\langle |f-c|, g\rangle.
\end{aligned}$$

Taking the supremum over all g with $\|g\|_r \le 1$ we get

$$\sum_e |df(e)|Q(e) \ge I_d \|f-c\|_q. \tag{3.3.8}$$

The desired inequality follows. Observe that in (3.3.8) c is a median of f.

PROOF OF THEOREM 3.3.11: Fix g and set $f = \operatorname{sgn}(g-c)|g-c|^2$ where c is a median of g, hence 0 is a median of f. The hypothesis of Theorem 3.3.11 implies that $I_d \ge 1/S$. Inequality (3.3.8) then shows that

$$\|g-c\|_{2q}^2 = \|f\|_q \le S \sum_e |df(e)Q(e)|.$$

As in the proof of Lemma 3.3.7 we have

$$\sum_e |df(e)|Q(e) \leq [8\,\mathcal{E}(g,g)]^{1/2}\|g-c\|_2.$$

Hence

$$\|g-c\|_{2q}^2 \leq [8S^2\,\mathcal{E}(g,g)]^{1/2}\|g-c\|_2.$$

Now, the Hölder inequality $\|h\|_2 \leq \|h\|_1^{1/(1+d)}\|h\|_{2q}^{d/(1+d)}$ yields

$$\|g-c\|_2 \leq \left([8S^2\,\mathcal{E}(f,f)]^{1/2}\|g-c\|_2\right)^{d/2(1+d)}\|g-c\|_1^{1/(1+d)}.$$

Thus

$$\|g-c\|_2^{2(1+2/d)} \leq 8S^2\mathcal{E}(f,f)\|g-c\|_1^{4/d}.$$

Since c is a median of g, it follows that

$$\mathrm{Var}_\pi(g)^{1+2/d} \leq 8S^2\,\mathcal{E}(f,f)\|g\|_1^{4/d}.$$

This is the desired result.

EXAMPLE 3.3.3: Consider a square grid $\mathcal{X} = \{0,\dots,n\}^2$ as in Theorem 3.3.8. The argument developed for Theorem 3.3.8 also yields the following result.

Theorem 3.3.14 *For the natural walk on the square grid $\mathcal{X} = \{0,\dots,n\}^2$ the isoperimetric constant I_2 (i.e., $d=2$) is given by*

$$I_2 = \begin{cases} \frac{1}{2^{3/2}(n+1)} & \text{if } n+1 \text{ is even} \\ \frac{(n+2)^{1/2}}{2^{3/2}n^{1/2}(n+1)} & \text{if } n+1 \text{ is odd.} \end{cases}$$

By Theorem 3.3.11 it follows that, for all $f \in \ell^2(\pi)$,

$$\mathrm{Var}_\pi(f)^2 \leq 64(n+1)^2\mathcal{E}(f,f)\|f\|_1^2.$$

By Theorem 2.3.2 this yields

$$\|h_t^x - 1\|_2 \leq \min\left\{2^{3/2}(n+1)/t^{1/2}, e^{-[t/64(n+1)^2]+1/2}\right\}.$$

This is a very good bound which is of the right order of magnitude **for all** $t>0$.

EXAMPLE 3.3.4: We can also compute I_d for a paralellepiped in d-dimensions.

Theorem 3.3.15 *For the natural walk on the parallelepiped*

$$\mathcal{X} = \{0,\dots,n_1\} \times \dots \times \{0,\dots,n_d\}$$

with $n_i \leq n_1$, the isoperimetric constant I_d satisfies

$$I_d \geq \frac{1}{d2^{1-1/d}(n_1+1)}$$

with equality if n_1+1 is even. It follows that

$$\mathrm{Var}_\pi(f)^{1+2/d} \leq 8 2^{2(1-1/d)}\, d^2\,(n_1+1)^2\,\mathcal{E}(f,f)\|f\|_1^{4/d}.$$

In [28] a somewhat better Nash inequality

$$\|f\|_2^{1+2/d} \le 64\, d\,(n+1)^2 \left(\mathcal{E}(f,f) + \frac{8}{d(n+1)^2}\|f\|_2^2\right)\|f\|_1^{4/d}$$

is proved (in the case $n_1 = \ldots = n_d = n$) by a different argument.

EXAMPLE 3.3.5: We now return to the "n-dog". The Nash inequality in Theorem 3.3.14 yields

$$\begin{aligned}\|f\|_2^2 &\le \left(64(n+1)^2\mathcal{E}(f,f)\|f\|_1^2\right)^{1/2} + \pi(f)^2 \\ &\le \left(64(n+1)^2\left(\mathcal{E}(f,f) + \frac{1}{64(n+1)^2}\|f\|_2^2\right)\|f\|_1^2\right)^{1/2}.\end{aligned}$$

for all functions f on a square grid $\{0,\ldots,n\}^2$. Now the n-dog is simply two square grids with one corner in common. Hence, applying the above inequality on each square grid, we obtain (the constant factor between the uniform distribution on one grid and the uniform distribution on the n-dog cancel)

$$\|f\|_2^4 \le 128(n+1)^2\left(\mathcal{E}(f,f) + \frac{1}{32(n+1)^2}\|f\|_2^2\right)\|f\|_1^2.$$

The change by a factor 2 in the numerical constants is due to the fact that the common corner o appears in each square grid. Recall that using Theorem 3.2.3 we have proved that the spectral gap of the dog is bounded below by

$$\lambda \ge \frac{1}{8(n+1)^2\log(2n+1)}.$$

Applying Theorem 2.3.5 and Corollary 2.3.5, we obtain the following result.

Theorem 3.3.16 *For the n-dog made of two square grids $\{0,\ldots,n\}^2$ with the corners $o = o_1 = o_2 = (0,0)$ identified, the natural chain satisfies*

$$\forall\, t \le 32(n+1)^2, \quad \|h_t^x\|_2 \le 8e(n+1)/t^{1/2}.$$

Also, for all $c > 0$ and $t = 8(n+1)^2(5 + c\log(2n+1))$

$$\|h_t^x - 1\|_2 \le e^{1-c}.$$

This shows that a time of order $n^2\log n$ suffices to reach stationarity on the n-dog. Furthermore, the upper bound on λ that we obtained earlier shows that this is optimal since $\max_x \|h_t^x - 1\|_1 \ge e^{-t\lambda} \ge e^{-at/(n^2\log n)}$.

Consider now all the eigenvalues $1 = \lambda_0 < \lambda_1 \le \ldots \le \lambda_{|\mathcal{X}|-1}$ of this chain. Corrolary 2.3.9 and Theorem 3.3.16 show that

$$\lambda_i \ge 10^{-4}(i+1)n^{-2}$$

for all $i \geq 10^4$. This is a good estimate except for the numerical constant 10^4. However, it leaves open the following natural question. We know that $\lambda = \lambda_1$ is of order $1/(n^2 \log n)$. How many eigenvalues are there such that $n^2 \lambda_i$ tends to zero as n tends to infinity? Interestingly enough the answer is that λ_1 is the only such eigenvalue. Namely, there exists a constant $c > 0$ such that, for $i \geq 2$, $\lambda_i \geq cn^{-2}$. We now prove this fact. Consider the squares

$$\mathcal{X}_- = \{-n, \dots, 0\}^2, \quad \mathcal{X}_+ = \{0, \cdots, n\}^2$$

and set

$$\psi_\pm(x) = \mathbf{1}_{\mathcal{X}_\pm}(x), \;\; x \in \mathcal{X}.$$

These functions span a two-dimensional vector space $E \subset \ell^2(\mathcal{X})$. On each of the two squares $\mathcal{X}_-, \mathcal{X}_+$, we have the Poincaré inequality

$$\sum_{x \in \mathcal{X}_\pm} |f(x)|^2 \leq \frac{1}{4}(n+1)^2 \sum_e |df(e)|^2 \tag{3.3.9}$$

for all function f on $\mathcal{X}_\pm$ satisfying $\sum_{x \in \mathcal{X}_\pm} f(x) = 0$. In this inequality, the right most sum runs over all edge e of the grid $\mathcal{X}_\pm$. There are many ways to prove this inequality. For instance, one can use Theorem 3.2.1 (with paths having only one turn), or the fact that the spectral gap is exactly $1 - \cos(\pi/(n+1))$ for the square grid.

Now, if f is a function in $\ell^2(\mathcal{X})$ which is orthogonal to E (i.e., to ψ_- and ψ_+), we can apply (3.3.9) to the restrictions f_+, f_- of f to $\mathcal{X}_+$, $\mathcal{X}_-$. Adding up the two inequalities so obtained we get

$$\forall \;\; f \in E^\perp, \quad \sum_{x \in \mathcal{X}} |f(x)|^2 \pi(x) \leq 2(n+1)^2 \mathcal{E}(f,f).$$

By the min-max principle (1.3.7), this shows that

$$\lambda_2 \geq \frac{1}{2(n+1)^2}.$$

Let ψ_1 denote the normalized eigenfunction associated to the spectral gap λ. For each n, let $a_n < b_n$ be such that

$$\lim_{n\to\infty} a_n n^{-2} = +\infty, \quad \lim_{n\to\infty} b_n [n^2 \log n]^{-1} = 0, \quad \lim_{n\to\infty} (b_n - a_n) = +\infty$$

and set $I_n = [a_n, b_n]$. Using the estimates obtained above for λ_1 and λ_2 together with Lemma 1.4.3 we conclude that for $t \in I_n$ and n large enough the density $h_t(x,y)$ of the semigroup H_t on the n-dog is close to

$$1 + \psi_1(x)\psi_1(y).$$

In words, the n-dog presents a sort of metastability phenomenon.

We finish this subsection by stating a bound on higher eigenvalues in terms of isoperimetry. It follows readily from Theorems 3.3.11 and 2.3.9.

Theorem 3.3.17 *Assume that* (K, π) *is reversible and satisfies* (3.3.7)*, that is,*

$$\pi(A)^{(d-1)/d} \leq SQ(\partial A)$$

for all $A \subset \mathcal{X}$ *such that* $\pi(A) \leq 1/2$*. Then the eigenvalues* λ_i *satisfy*

$$\lambda_i \geq \frac{i^{2/d}}{8e^{2/d}dS^2}.$$

Compare with [14].

3.3.3 Isoperimetry and the log-Sobolev constant

Theorem 2.3.6 can be used, together with theorems 3.3.10, 3.3.11, to bound the log-Sobolev constant α from below in terms of isoperimetry. This yields the following results.

Theorem 3.3.18 *Let* (K, π) *be a finite reversible Markov chain.*

1. *Assume* (K, π) *satisfies* (3.3.7)*, that is,*

   $$\pi(A)^{(d-1)/d} \leq SQ(\partial A)$$

   for all $A \subset \mathcal{X}$ *such that* $\pi(A) \leq 1/2$*. Then the log-Sobolev constant* α *is bounded below by*

   $$\alpha \geq \frac{1}{4dS^2}.$$

2. *Assume instead that* (K, π) *satisfies* (3.3.4)*, that is,*

   $$\pi(A)^{(d-1)/d} \leq S\left(Q(\partial A) + \frac{1}{R}\pi(A)\right),$$

   for all set $A \subset \mathcal{X}$*. Then*

   $$\alpha \geq \frac{\lambda}{2\left[1 + 8R^2\lambda + \frac{d}{4}\log\left(\frac{dS^2}{2R^2}\right)\right]}.$$

EXAMPLE 3.3.6: Theorem 3.3.18 and Theorems 3.3.14, 3.3.16 prove that the two-dimensional square grid $\mathcal{X} = \{0, \ldots, n\}^2$ or the two-dimensional n-dog have $\alpha \simeq \lambda$. Namely, for the two-dimensional n-grid, α and λ are of order $1/n^2$ whereas, for the n-dog, α and λ are of order $1/[n^2 \log n]$.

EXAMPLE 3.3.7: For the d-dimensional square grid $\mathcal{X} = \{0, \ldots, n\}^d$, applying Theorems 3.3.18 and 3.3.15 we obtain

$$\alpha \geq \frac{2}{d^3(n+1)^2}$$

whereas Lemma 2.2.11 can be used to show that α is of order $1/[dn^2]$ in this case.

3.4 Moderate growth

This section presents geometric conditions that implies that a Nash inequality holds. More details and many examples can be found in [25, 26, 28]. Let us emphasize that the notions of **moderate growth** and of **local Poincaré inequality** presented briefly below are really instrumental in proving useful Nash inequalities in explicit examples. See [28].

Definition 3.4.1 *Let (K, π) be an irreducible Markov chain on a finite state space $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge set according to Definition* 3.1.1. *Let $d(x, y)$ denote the distance between x and y in $(\mathcal{X}, \mathcal{A})$ and $\gamma = \max_{x,y} d(x, y)$ be the diameter. Define*

$$V(x, r) = \pi(\{y : d(x, y) \le r\}).$$

(1) We say the (K, π) has (M, d)-moderate growth if

$$V(x, r) \ge \frac{1}{M}\left(\frac{r+1}{\gamma}\right)^d \quad \textit{for all } x \in \mathcal{X} \textit{ and all } r \le \gamma.$$

(2) We say that (K, π) satisfies a local Poincaré inequality with constant $a > 0$ if

$$\|f - f_r\|_2^2 \le ar^2\mathcal{E}(f, f) \quad \textit{for all functions } f \textit{ and all } r \le \gamma$$

where

$$f_r(x) = \frac{1}{V(x, r)} \sum_{y:d(x,y)\le r} f(y)\pi(y).$$

Moderate growth is a purely geometric condition. On one hand it implies (take $r = 0$) that $\pi_* \ge M^{-1}\gamma^{-d}$. If π is uniform, this says $|\mathcal{X}| \le M\gamma^d$. On the other hand, it implies that the volume of a ball of radius r grows at least like r^d.

The local Poincaré inequality implies in particular (take $r = \gamma$) that $\mathrm{Var}_\pi(f) \le a\gamma^2\mathcal{E}(f, f)$, that is $\lambda \ge 1/(a\gamma^2)$. It can sometimes be checked using the following lemma.

Lemma 3.4.2 *For each $(x, y) \in \mathcal{X}^2$, $x \ne y$, fix a path $\gamma(x, y)$ in $\Gamma(x, y)$. Then*

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

where

$$\eta(r) = \max_{e\in\mathcal{A}} \left\{ \frac{2}{Q(e)} \sum_{\substack{x,y:d(x,y)\le r,\\ \gamma(x,y)\ni e}} |\gamma(x, y)| \frac{\pi(x)\pi(y)}{V(x, r)} \right\}.$$

See [28], Lemma 5.1.

Definition 3.4.1 is justified by the following theorem.

Theorem 3.4.3 *Assume that (K,π) has (M,d) moderate growth and satisfies a local Poincaré inequality with constant $a>0$. Then $\lambda \geq 1/a\gamma^2$ and (K,π) satisfies the Nash inequality*

$$\|f\|_2^{2(1+2/d)} \leq C\left(\mathcal{E}(f,f)+\frac{1}{a\gamma^2}\|f\|_2^2\right)\|f\|_1^{4/d}$$

with $C=(1+1/d)^2(1+d)^{2/d}M^{2/d}a\gamma^2$. It follows that

$$\|h_t^x-1\|_2 \leq Be^{-c} \quad \text{for } t=a\gamma^2(1+c),\quad c>0$$

with $B=(e(1+d)M)^{1/2}(2+d)^{d/4}$. Also, the log-Sobolev constant satisfies $\alpha \geq \varepsilon/\gamma^2$ with $\varepsilon^{-1}=2a(2+\log B)$.

Futhermore, there exist constants c_i, $i=1,\dots,6$, depending only on M,d,a and such that $\lambda \leq c_1/\gamma^2$, $\alpha \leq c_2/\gamma^2$ and, if (K,π) is reversible,

$$c_3e^{-c_4t/\gamma^2} \leq \max_x \|h_t^x-1\|_1 \leq c_5e^{-c_6t/\gamma^2}.$$

See [28], Theorems 5.2, 5.3 and [29], Theorem 4.1.

One can also state the following result for higher eigenvalues of reversible Markov chains.

Theorem 3.4.4 *Assume that (K,π) is reversible, has (M,d) moderate growth and satisfies a local Poincaré inequality with constant $a>0$. Then there exists a constant $c=c(M,d,a)>0$ such that $\lambda_i \geq ci^{2/d}\gamma^{-2}$.*

Chapter 4

Comparison techniques

This chapter develops the idea of comparison between two finite chains K, K'. Typically we are interested in studying a certain chain K on $\mathcal{X}$. We consider an auxilliary chain K' on $\mathcal{X}$ or even on a different but related state space $\mathcal{X}'$. This auxilliary chain is assumed to be well-known, and the chain K is not too different from K'. Comparison techniques allow us to transfer information from K to K'. We have already encounter this idea several times. It is emphasized and presented in detail in this chapter. The main references for this chapter are [23, 24, 30].

4.1 Using comparison inequalities

This section collects a number of results that are the keys of comparison techniques. Most of these results have already been proved in previous chapters, sometimes under less restrictive hypoheses.

Theorem 4.1.1 *Let (K, π), (K', π') be two irreducible finite chains defined on two state spaces $\mathcal{X}$, $\mathcal{X}'$ with $\mathcal{X} \subset \mathcal{X}'$. Assume that there exists an extention map $f \to \tilde{f}$ that associates a function $\tilde{f} : \mathcal{X} \to \mathbb{R}$ to any function $f : \mathcal{X}' \to \mathbb{R}$ and such that $\tilde{f}(x) = f(x)$ if $x \in \mathcal{X}$. Assume further that there exist $a, A > 0$ such that*

$$\forall\, f : \mathcal{X} \to \mathbb{R}, \quad \mathcal{E}'(\tilde{f}, \tilde{f}) \le A\mathcal{E}(f, f) \quad \textit{and} \quad \forall\, x \in \mathcal{X}, \;\; a\pi(x) \le \pi'(x).$$

Then

(1) The spectral gaps λ, λ' and the log-Sobolev constants α, α' satisfy

$$\lambda \ge a\lambda'/A\,, \quad \alpha \ge a\alpha'/A.$$

In particular

$$\|h_t^x - 1\|_2 \le e^{1-c} \quad \textit{for all} \quad t = \frac{Ac}{a\lambda'} + \frac{A}{2a\alpha'} \log_+ \log \frac{1}{\pi(x)} \;\; \textit{with } c > 0.$$

(2) If (K,π) and (K',π') are reversible chains, and $|\mathcal{X}| = n$, $|\mathcal{X}'| = n'$,

$$\forall\, i = 1,\dots,n-1, \quad \lambda_i \geq a\lambda_i'/A$$

where $(\lambda_i)_0^{n-1}$ (resp $(\lambda_i')_0^{n'-1}$) are the eigenvalues of $I-K$ (resp. $I-K'$) in nondecreasing order. In particular, for all $t>0$,

$$|\!|\!|h_t - 1|\!|\!|^2 \quad \leq \quad |\!|\!|h'_{at/A} - 1|\!|\!|^2 = \sum_1^{n'-1} e^{-2at\lambda_i'/A}$$

where

$$|\!|\!|h_t - 1|\!|\!|^2 = \sum_{x,y} |h_t(x,y)-1|^2\pi(x)\pi(y) = \sum_x \|h_t^x - 1\|_2^2\pi(x).$$

(3) If (K,π) and (K',π') are reversible chains and that there exists a group G that acts transitively on $\mathcal{X}$ with $K(gx,gy) = K(x,y)$ and $\pi(gx) = \pi(x)$ then

$$\forall\, x \in \mathcal{X}, \quad \|h_t^x - 1\|_2^2 \leq \sum_1^{n'-1} e^{-2at\lambda_i'/A}.$$

(4) If (K,π) and (K',π') are invariant under transitive group actions then

$$\forall\, x \in \mathcal{X},\ x' \in \mathcal{X}', \quad \|h_t^x - 1\|_2 \leq \|h'^{x'}_{at/A} - 1\|_2.$$

Proof: The first assertion follows from Lemma 2.2.12 and Corollary 2.2.4. The second uses Theorem 1.3.4 and (1.3.6). The last statement simply follows from *(2)* and the fact that $\|h_t^x - 1\|_2$ does not depend on x under the hypotheses of *(3)*. Observe that the theorem applies when $\mathcal{X} = \mathcal{X}'$. In this case the extention map $f \to \tilde{f} = f$ is the identity map on functions.

These results shows how the comparison of the Dirichlet forms $\mathcal{E}$, $\mathcal{E}'$ allows us to bound the convergence of h_t towards π in terms of certain parameters related to the chain K' which we assume we understand better. The next example illustrates well this technique.

Example 4.1.1: Let $\mathcal{Z} = \{0,1\}^n$. Fix a nonnegative sequence $\mathbf{a} = (a_i)_1^n$ and $b \geq 0$. Set

$$\mathcal{X}(\mathbf{a},b) = \mathcal{X} = \left\{x = (x_i)_1^n \in \mathcal{Z} : \sum a_i x_i \leq b\right\}.$$

On this set, consider the Markov chain with Kernel

$$K_{\mathbf{a},b}(x,y) = K(x,y) = \begin{cases} 0 & \text{if } |x-y|>1 \\ 1/n & \text{if } |x-y| = 1 \\ (n-n(x))/n & \text{if } x = y \end{cases}$$

where $n(x) = n_{\mathbf{a},b}(x)$ is the number of $y \in \mathcal{X}$ such that $|x-y| = 1$, that is, the number of neighbors of x in $\mathcal{Z}$ that are in $\mathcal{X}$. Observe that this definition makes sense for any (say connected) subset of $\mathcal{Z}$. This chains is symmetric and has the uniform distibution $\pi \equiv 1/|\mathcal{X}|$ as reversible measure.

For instance, in the simple case where $a_i = 1$ for all i,

$$\mathcal{X}(\mathbf{1}, b) = \left\{ x \in \{0,1\}^n : \sum_i x_i \le b \right\}$$

and

$$K_{\mathbf{1},b}(x,y) = \begin{cases} 0 & \text{if } |x-y| > 1 \\ 1/n & \text{if } |x-y| = 1 \\ (n-b)/n & \text{if } x = y \text{ and } |x| = b. \end{cases}$$

As mentioned in the introduction, proving that a polynomial time $t = O(n^A)$ suffices to insure convergence of this chain, uniformly over all possible choices of $\mathbf{a}, b$, is an open problem.

Here we will prove a partial result for $\mathbf{a}, b$ such that $\mathcal{X}(\mathbf{a}, b)$ is big enough. Set $|x| = \sum_1^n x_i$. Set also $x \le y$ (resp. $<$) if $x_i \le y_i$ (resp. $<$) for $x, y \in \mathcal{Z}$. Clearly, $y \in \mathcal{X}(\mathbf{a}, b)$ and $x \le y$ implies that $x \in \mathcal{X}(\mathbf{a}, b)$. Furthermore, if $|x - y| = 1$, then either $x < y$ or $y < x$. Set

$$V^{\downarrow}(x) = \{y \in \mathcal{Z} : |x-y| = 1, y < x\}.$$

Now, we fix $\mathbf{a} = (a_i)_1^n$ and b. For each integer c let $\mathcal{X}_c$ be the set

$$\mathcal{X}_c = \mathcal{X} \bigcup \left\{ z \in \mathcal{Z} : \sum x_i \le c \right\}.$$

Hence $\mathcal{X}_{c+1}$ is obtained from $\mathcal{X}_c$ by adding the points z with $\sum z_i = c+1$. On each $\mathcal{X}_c$ we consider the natural chain defined as above. We denote by

$$\mathcal{E}_c(f,f) = \frac{1}{2n|\mathcal{X}^c|} \sum_{\substack{x,y \in \mathcal{X}^c \\ |x-y|=1}} |f(x) - f(y)|^2$$

its Dirichlet form. We will also use the notation π_c, Var_c, λ_c, α_c.

Define ℓ to be the largest integer such that $\sum_{i \in I} a_i \le b$ for all subsets $I \subset \{1, \dots, n\}$ with $\#I = \ell$. Observe that $\mathcal{X}_c = \mathcal{X}$ for $c \le \ell$. Also, $\mathcal{X}_n = \mathcal{Z} = \{0,1\}^n$. We claim that the following inequalities hold between the spectral gaps and log-Sobolev constants of the natural chains on $\mathcal{X}^c$, $\mathcal{X}^{c+1}$.

$$\lambda_{c+1} \le \left(1 + \frac{2(n-c)}{c+1}\right) \lambda_c \tag{4.1.1}$$

$$\alpha_{c+1} \le \left(1 + \frac{2(n-c)}{c+1}\right) \alpha_c. \tag{4.1.2}$$

If we can prove these inequalities, it will follow that

$$\frac{2}{n} \le e^{2\frac{(n-\ell)^2}{\ell+1}} \lambda(\mathbf{a}, b) \tag{4.1.3}$$

$$\frac{1}{n} \le e^{2\frac{(n-\ell)^2}{\ell+1}} \alpha(\mathbf{a}, b) \tag{4.1.4}$$

where $\lambda(\mathbf{a}, b)$ and $\alpha(\mathbf{a}, b)$ are the spectral gap and log-Sobolev constant of the chain $K = K_{\mathbf{a},b}$ on $\mathcal{X} = \mathcal{X}_{\mathbf{a},b}$. To see this use

$$\sum_{c=\ell}^{n-1} \frac{n-c}{c+1} \le (n-\ell) \sum_{\ell}^{n-1} \frac{1}{c+1} \le \frac{(n-\ell)^2}{\ell+1}.$$

To prove (4.1.1), (4.1.2) we proceed as follows. Fix $c \ge \ell$. Given a function $f : \mathcal{X}_c \to \mathbb{R}$ we extend it to a function $\tilde{f} : \mathcal{X}_{c+1} \to \mathbb{R}$ by the formula

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \mathcal{X}^c \\ \frac{1}{c+1} \sum_{y \in V^{\downarrow}(x)} f(y) & \text{if } x \in \mathcal{X}_{c+1} \setminus \mathcal{X}_c \end{cases}$$

(observe that $\#V^{\downarrow}(x) = c+1$ if $|x| = c+1$). With this definition, we have

$$\begin{aligned} \mathrm{Var}_c(f) &\le \sum_{x \in \mathcal{X}_c} |f(x) - \pi_{c+1}(\tilde{f})|^2 \frac{1}{|\mathcal{X}_c|} \\ &\le \frac{|\mathcal{X}_{c+1}|}{|\mathcal{X}_c|} \sum_{x \in \mathcal{X}_{c+1}} |\tilde{f}(x) - \pi_{c+1}(\tilde{f})|^2 \frac{1}{|\mathcal{X}_{c+1}|} \le \frac{|\mathcal{X}_{c+1}|}{|\mathcal{X}_c|} \mathrm{Var}_{c+1}(\tilde{f}) \end{aligned}$$

and, similarly, $\mathcal{L}_c(f) \le [|\mathcal{X}_{c+1}|/|\mathcal{X}_c|]\, \mathcal{L}_{c+1}(\tilde{f})$. We can also bound $\mathcal{E}_{c+1}(\tilde{f}, \tilde{f})$ in terms of $\mathcal{E}_c(f, f)$.

$$\begin{aligned} \mathcal{E}_{c+1}(\tilde{f}, \tilde{f}) &= \frac{1}{2n|\mathcal{X}_{c+1}|} \sum_{\substack{x,y \in \mathcal{X}_{c+1}: \\ |x-y|=1}} |\tilde{f}(x) - \tilde{f}(y)|^2 \\ &\le \frac{|\mathcal{X}_c|}{|\mathcal{X}_{c+1}|} \left(\frac{1}{2n|\mathcal{X}_c|} \sum_{\substack{x,y \in \mathcal{X}_c: \\ |x-y|=1}} |f(x) - f(y)|^2 \right. \\ &\qquad \left. + \frac{1}{n|\mathcal{X}_c|} \sum_{x:|x|=c+1} \sum_{y \in V^{\downarrow}(x)} |\tilde{f}(x) - f(y)|^2 \right) \\ &= \frac{|\mathcal{X}_c|}{|\mathcal{X}_{c+1}|} \left(\mathcal{E}_c(f,f) + \frac{1}{n|\mathcal{X}_c|} \mathcal{R} \right). \end{aligned}$$

We now bound $\mathcal{R}$ in terms of $\mathcal{E}_c(f, f)$. If $|x - y| = 2$, let $x \wedge y$ be the unique element in $V^{\downarrow}(x) \cap V^{\downarrow}(y)$.

$$\begin{aligned} \mathcal{R} &= \sum_{x:|x|=c+1} \sum_{y \in V^{\downarrow}(x)} |\tilde{f}(x) - f(y)|^2 \\ &= \sum_{x:|x|=c+1} \frac{1}{2(c+1)} \sum_{y,z \in V^{\downarrow}(x)} |f(z) - f(y)|^2 \\ &\le \sum_{x:|x|=c+1} \frac{1}{c+1} |f(z) - f(z \wedge y)|^2 + |f(z \wedge y) - f(y)|^2 \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{x:|x|=c+1} \frac{2}{c+1} \sum_{\substack{v\in V^{\downarrow}(x)\\ u\in V^{\downarrow}(v)}} |f(v)-f(u)|^2 \\
&\leq \frac{2n(n-c)|\mathcal{X}_c|}{c+1}\mathcal{E}_c(f,f).
\end{aligned}$$

Hence

$$\mathcal{E}_{c+1}(\tilde f,\tilde f) \leq \frac{|\mathcal{X}_c|}{|\mathcal{X}_{c+1}|}\left(1+\frac{2(n-c)}{c+1}\right)\mathcal{E}_c(f,f).$$

Now, Lemmma 2.2.12 yields the claimed inequalities (4.1.1) and (4.1.2). We have proved the following result.

Theorem 4.1.2 *Assume that* $\mathbf{a}=(a_i)_1^n$*,* b *and* ℓ *are such that* $a_i, b\geq 0$ *and* $\sum_{i\in I} a_i \leq b$ *for all* $I\subset\{1,\dots,n\}$ *satisfying* $\#I \leq n-n^{1/2}$*. Then the chain* $K_{\mathbf{a},b}$ *on*

$$\mathcal{X}(\mathbf{a},b)=\left\{x=(x_i)_1^n\in\{0,1\}^n : \sum_i a_i x_i \leq b\right\}$$

satisfies

$$\lambda(\mathbf{a},b)\geq\frac{2\epsilon}{n},\quad \alpha(\mathbf{a},b)\geq\frac{\epsilon}{n}\quad \text{with}\quad \epsilon=e^{-4}$$

The associated semigroup $H_t=H_{\mathbf{a},b,t}=e^{-t(I-K_{\mathbf{a},b})}$ *satisfies*

$$\|h_t^x-1\|_2\leq e^{1-c}\ \text{for}\ t=(4\epsilon)^{-1}n\left(\log n\ +2c\right).$$

These are good estimates and I believe it would be difficult to prove similar bounds for $\|h_t^x-1\|_2$ without using the notion of log-Sobolev constant (coupling is a possible candidate but if it works, it would only give a bound in ℓ^1).

In the case where $a_i=1$ for all i and $b\geq n/2$, we can use the test function $f(x)=\sum_{i<n/2}(x_i-1/2)-\sum_{i>n/2}(x_i-1/2)$ to bound $\lambda(\mathbf{1},b)$ and $\alpha(\mathbf{1},b)$ from above. Indeed, this function satisfies $\pi_{\mathbf{1},b}(f)=\pi_{\mathcal{Z}}(f)=0$ (use the symmetry that switches $i<n/2$ and $i>n/2$) and $\mathrm{Var}_{\mathbf{1},b}(f,f)\geq 2\frac{|\mathcal{Z}|}{|\mathcal{X}(\mathbf{a},b)|}\mathrm{Var}_{\mathcal{Z}}(f,f)$ (use the symmetry $x\to x+\mathbf{1}$ mod (2)). Also $\mathcal{E}_{\mathbf{a},b}\leq\frac{|\mathcal{X}(\mathbf{a},b)|}{|\mathcal{Z}|}\mathcal{E}_{\mathcal{Z}}$. Hence $\lambda(\mathbf{a},b)\leq 4/n$, $\alpha(\mathbf{a},b)\leq 2/n$ in this particular case.

4.2 Comparison of Dirichlet forms using paths

The path technique of Section 3.1 can be used to compare two Dirichlet forms on a same state space $\mathcal{X}$. Together with Theorem 4.1.1 this provides a powerful tool to study finite Markov chains that are not too different from a given well-known chain. The results presented below can be seen as extentions of Theorems 3.2.1, 3.2.5. Indeed, what has been done in these theorems is nothing else than comparing the chain (K,π) of interest to the "trivial" chain with kernel $K'(x,y)=\pi(y)$ which has the same stationary distribution π. This chain K' has Dirichlet form

$\mathcal{E}'(f,f) = \mathrm{Var}_\pi(f)$ and is indeed well-known: It has eigenvalue 1 with multiplicity 1 and all the other eigenvalues vanish. Its log-Sobolev constant is given in Theorem 2.2.9. Once the Theorems of Section 3.2 have been interpreted in this manner their generalization presented below is straight-forward.

We will use the following notation. Let (K,π) be the unknown chain of interest and

$$Q(e) = \frac{1}{2}\left(K(x,y)\pi(x) + K(y,x)\pi(y)\right) \;\text{ if }\; e = (x,y).$$

Let $\mathcal{A}$ be an adapted edge-set according to Definition 3.1.1 and let

$$\Gamma = \bigcup_{x,y} \Gamma(x,y)$$

where $\Gamma(x,y)$ be the set of all paths from x to y that have no repeated edges.

Theorem 4.2.1 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge-set for K. Let (K',π') be an auxilliary chain. For each $(x,y) \in \mathcal{X}\times\mathcal{X}$ such that $x \neq y$ and $K'(x,y) > 0$ choose exactly one path $\gamma(x,y)$ in $\Gamma(x,y)$. Then $\mathcal{E}' \leq A\mathcal{E}$ where*

$$A = \max_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y\in\mathcal{X}:\\ \gamma(x,y)\ni e}} |\gamma(x,y)| K'(x,y)\pi'(x) \right\}.$$

PROOF: For each $(x,y) \in \mathcal{X}\times\mathcal{X}$ such that $K'(x,y) > 0$, write

$$f(y) - f(x) = \sum_{e\in\gamma(x,y)} df(e)$$

and, using Cauchy-Schwarz,

$$|f(y) - f(x)|^2 \leq |\gamma(x,y)| \sum_{e\in\gamma(x,y)} |df(e)|^2.$$

Multiply by $\frac{1}{2}K'(x,y)\pi'(x)$ and sum over all x,y to obtain

$$\frac{1}{2}\sum_{x,y} |f(y)-f(x)|^2 K'(x,y)\pi'(x) \leq \frac{1}{2}\sum_{x,y} |\gamma(x,y)| \sum_{e\in\gamma(x,y)} |df(e)|^2 K'(x,y)\pi(x).$$

The left-hand side is equal to $\mathcal{E}'(f,f)$ whereas the right-hand side becomes

$$\frac{1}{2}\sum_{e\in\mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{x,y:\\ \gamma(x,y)\ni e}} |\gamma(x,y)K'(x,y)|\pi'(x) \right\} |df(e)|^2 Q(e)$$

which is bounded by

$$\max_{e\in\mathcal{A}}\left\{\frac{1}{Q(e)}\sum_{\substack{x,y:\\ \gamma(x,y)\ni e}}|\gamma(x,y)|K'(x,y)\pi'(x)\right\}\mathcal{E}(f,f).$$

Hence

$$\forall f,\quad \mathcal{E}(f,f)\le A\mathcal{E}(f,f)$$

with A as in Theorem 4.2.1.

Theorems 4.1.1, 4.2.1 are helpful for two reasons. First, non-trivial informations about K' can be brought to bear in the study of K. Second, the path combinatorics that is involved in Theorem 4.2.1 is often simpler than that involved in Theorem 3.2.1 because only the pairs (x,y) such that $K'(x,y)>0$ enter in the bound. These two points are illustrated by the next example.

EXAMPLE 4.2.1: Let $\mathcal{X}=\{0,1\}^n$. Let $x\to\tau(x)$, be defined by $[\tau(x)]_i=x_{i-1}$, $1<i\le n$, $[\tau(x)]_1=x_n$. Let $x\to\sigma(x)$ be defined by $\sigma(x)=x+(1,0,\dots,0)$. Set $K(x,y)=1/n$ if either $y=\tau^j(x)$ for some $1\le j\le n$ or $y=\sigma(x)$, and $K(x,y)=0$ otherwise. This chain is reversible with respect to the uniform distribution. In Section 3.2, we have seen that $\lambda\ge 1/n^3$ by Theorem 3.2.1. Here, we compare K with the chain $K'(x,y)=1/n$ if $|x-y|=1$ and $K(x,y)=0$ otherwise. For (x,y) with $|x-y|=1$, let i be such that $x_i\ne y_i$. Let

$$\gamma(x,y)=(x,\tau^j(x),\sigma\circ\tau^j(x),\tau^{-j}\circ\sigma\circ\tau^j(x)=y)$$

where $j=i$ if $i\le n/2$ and $j=n-i$ if $i>n/2$. These paths have length 3. The constant A of Theorem 4.2.1 becomes

$$A=3\max_{e\in\mathcal{A}}\#\left\{(x,y):K'(x,y)>0,\ \gamma(x,y)\ni e\right\}.$$

If $e=(u,v)$ with $v=\tau^j(u)$, there are only two (x,y) such that $e\in\gamma(x,y)$ depending on whether σ appears after or before e. If $v=\sigma(u)$, there are n possibilities depending on the choice of $j\in\{0,1,\dots,n-1\}$. Hence $A=3n$. Since $\lambda'=2/n$ and $\alpha'=1/n$, this yields

$$\lambda\ge\frac{2}{3n^2},\qquad \alpha\ge\frac{1}{3n^2}.$$

Also it follows that

$$\max_x\|h_t^x-1\|_2\le e^{1-c}\quad\text{for}\quad t=\frac{3n^2}{4}\left(2c+\log n\right),\ \ c>0.$$

EXAMPLE 4.2.2: Consider a graph $(\mathcal{X},\mathcal{A})$ where $\mathcal{A}$ is a symmetric set of oriented edges. Set $d(x)=\#\{y\in\mathcal{X}:(x,y)\in\mathcal{A}\}$ and

$$K(x,y)=\left\{\begin{array}{cl}0 & \text{if } (x,y)\notin\mathcal{A}\\ 1/d(x) & \text{if } (x,y)\in\mathcal{A}.\end{array}\right.$$

This is the kernel of the simple random walk on $(\mathcal{X}, \mathcal{A})$. It is reversible with respect to the measure $\pi(x) = d(x)/|\mathcal{A}|$. For each $(x, y) \in \mathcal{X}^2$ choose a path $\gamma(x, y)$ with no repeated edges. Set

$$d_* = \max_{x \in \mathcal{X}} d(x), \quad \gamma_* = \max_{x,y \in \mathcal{X}} |\gamma(x, y)|, \quad \eta_* = \max_{e \in \mathcal{A}} \#\{(x, y) \in \mathcal{X}^2 : \gamma(x, y) \ni e\}.$$

We now compare with the chain $K'(x, y) = 1/|\mathcal{X}|$ which has reversible measure $\pi'(x) = 1/|\mathcal{X}|$ and spectral gap $\lambda' = 1$. Theorem 4.2.1 gives $\lambda \geq a/A$ with

$$A \leq \frac{|\mathcal{A}|\gamma_*\eta_*}{|\mathcal{X}|^2} \quad \text{and} \quad a = \frac{|\mathcal{A}|}{d_*|\mathcal{X}|}.$$

This gives

Theorem 4.2.2 *For the simple random walk on a graph $(\mathcal{X}, \mathcal{A})$ the spectral gap is bounded by*

$$\lambda \geq \frac{|\mathcal{X}|}{d_*\gamma_*\eta_*}.$$

Compare with Example 3.2.4 where we used Theorem 3.2.1 instead. The present result is slightly better than the bound obtained there. It is curious that one obtains a better bound by comparing with the chain $K'(x, y) = 1/|\mathcal{X}|$ as above than by comparing with the $\widetilde{K}(x, y) = \pi(y)$ which corresponds to Theorem 3.2.1.

It is a good exercise to specialize Theorem 4.2.1 to the case of two left invariant Markov chains $K(x, y) = q(x^{-1}y)$, $K'(x, y) = q'(x^{-1}y)$ on a finite group G. To take advantage of the group invariance, write any element g of G as a product

$$g = g_1^{\epsilon_1} \cdots g_k^{\epsilon_k}$$

with $q(g_i) + q(g_i^{-1}) > 0$. View this as a path $\gamma(g)$ from the identity id of G to g. Then for each (x, y) with $q'(x^{-1}y) > 0$, write

$$x^{-1}y = g(x, y) = g_1^{\epsilon_1} \cdots g_k^{\epsilon_k}$$

(where the g_i and ϵ_i depend on (x, y)) and define

$$\gamma(x, y) = x\gamma(g) = (x, xg_1, \ldots, xg_1 \ldots g_{k-1}, xg(x, y) = y).$$

With this choice of paths Theorem 4.2.1 yields

Theorem 4.2.3 *Let K, K' be two invariant Markov chains on a group G. Set $q(g) = K(\mathrm{id}, g)$, $q'(g) = K'(\mathrm{id}, g)$. Let π denote the uniform distribution. Fix a generating set S satisfying $S = S^{-1}$ and such that $q(s) + q(s^{-1}) > 0$. for all $s \in S$. For each $g \in G$ such that $q'(g) > 0$, choose a writing of g as a product of elements of S, $g = s_1 \ldots s_k$ and set $|g| = k$. Let $N(s, g)$ be the number of times $s \in S$ is used in the chosen writing of g. Then $\mathcal{E} \leq A\mathcal{E}'$ and $\lambda \geq \lambda'/A$ with*

$$A = \max_{s \in S} \left\{ \frac{2}{q(s) + q(s^{-1})} \sum_{g \in G} |g| N(s, g) q'(g) \right\}.$$

Assume further that K, K' are reversible and let λ_i (resp. λ'_i), $i = 0, \ldots, |G|-1$ denote the eigenvalues of $I-K$ (resp. $I-K'$) in non-decreasing order. Then $\lambda_i \geq \lambda'_i/A$ for all $i \in \{1, \ldots, |G|-1\}$ and

$$\forall\, x \in G, \quad \|h_t^x - 1\|_2 \leq \|h'^x_{t/A} - 1\|_2.$$

PROOF: (cf. [23], pg 702) We use Theorem 4.2.1 with the paths described above. Fix an edge $e = (z, w)$ with $w = zs$. Observe that there is a bijection between

$$\{(g,h) \in G \times G : \gamma(g,h) \ni (z,w)\}$$

and

$$\{(g,u) \in G \times G : \exists\, i \text{ such that } s_i(u) = s, z = gs_1(u)\cdots s_{i-1}(u)\}$$

given by $(g,h) \to (g, g^{-1}h) = (g,u)$. For each fixed $u = g^{-1}h$, there are exactly $N(s,u)$ $g \in G$ such that (g,u) belongs to

$$\{(x,u) \in G \times G : \exists\, i \text{ such that } s_i(u) = s, z = xxs_1(u)\cdots s_{i-1}(u)\}.$$

Hence

$$\sum_{(g,h) \in G \times G : \gamma(g,h) \ni (z,w)} |\gamma(g,h)| = \sum_{u \in G} |u| N(s,u).$$

This proves the desired result. See also [24] for a more direct argument.

We now extend Theorem 4.2.1 to allow the use of a set of paths for each pair (x,y) with $K'(x,y) > 0$.

Definition 4.2.4 *Let (K, π), K', π' be two irreducible Markov chains on a same finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge-set for (K, π). A (K, K')-flow is non-negative function $\phi : \Gamma(K') \to [0, \infty[$ on the path set*

$$\Gamma(K') = \bigcup_{\substack{x,y:\\ K'(x,y)>0}} \Gamma(x,y)$$

such that

$$\forall x, y \in \mathcal{X},\ x \neq y,\ K'(x,y) > 0, \quad \sum_{\gamma \in \Gamma(x,y)} \phi(\gamma) = K'(x,y)\pi'(x).$$

Theorem 4.2.5 *Let K be an irreducible chain with stationary measure π on a finite set $\mathcal{X}$. Let $\mathcal{A}$ be an adapted edge-set for (K, π). Let (K', π') be a second chain and ϕ be a (K, K')-flow. Then $\mathcal{E}' \leq A(\phi)\mathcal{E}$ where*

$$A(\phi) = \max_{e \in \mathcal{A}} \left\{ \frac{1}{Q(e)} \sum_{\substack{\gamma \in \Gamma(K'):\\ \gamma \ni e}} |\gamma| \phi(\gamma) \right\}.$$

PROOF: For each (x,y) such that $K'(x,y)>0$ and each $\gamma\in\Gamma(x,y)$ write

$$|f(y)-f(x)|^2 \le |\gamma| \sum_{e\in\gamma} |df(e)|^2.$$

Then

$$|f(y)-f(x)|^2 K'(x,y)\pi'(x) \le \sum_{\gamma\in\Gamma(x,y)} |\gamma| \sum_{e\in\gamma} |df(e)|^2 \phi(\gamma).$$

From here, complete the proof as for Theorem 4.2.1.

Corollary 4.2.6 *Assume that there is a group G which acts on $\mathcal{X}$ and such that*

$$\pi(gx)=\pi(x),\ \ \pi'(gx)=\pi'(x),\ \ Q(gx,gy)=Q(x,y),\ \ Q'(gx,gy)=Q'(x,y).$$

Let $\mathcal{A}$ be an adapted edge-set for (K,π) such that $(x,y)\in\mathcal{A}\Rightarrow(gx,gy)\in\mathcal{A}$. Let $\mathcal{A}=\bigcup_1^k\mathcal{A}_i$, be the partition of $\mathcal{A}$ into transitive classes for this action. Then $\mathcal{E}'\le A\mathcal{E}$ where

$$A=\max_{1\le i\le k}\left\{\frac{1}{|\mathcal{A}_i|Q_i}\sum_{x,y} N_i(x,y)d_K(x,y)K'(x,y)\pi(x)\right\}.$$

Here $|\mathcal{A}_i|=\#\mathcal{A}_i$, $Q_i=Q(e_i)$ with $e_i\in\mathcal{A}_i$, $d_K(x,y)$ is the distance between x and y in $(\mathcal{X},\mathcal{A})$, and $N_i(x,y)$ is the maximum number of edges of type i in a geodesic path from x to y.

PROOF: Consider the set $\mathcal{G}(x,y)$ of all geodesic paths from x to y. Define a (K,K')-flow ϕ by setting

$$\phi(\gamma)=\begin{cases} K'(x,y)\pi'(x)/\#\mathcal{G}(x,y) & \text{if } \gamma\in\mathcal{G}(x,y)\\ 0 & \text{otherwise.}\end{cases}$$

Then $A(\phi)=\max_e A(\phi,e)$ where

$$A(\phi,e)=\frac{1}{Q(e)}\sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}}|\gamma|\phi(\gamma).$$

By hypothesis, $A(\phi,e_i)=A_i(\phi)$ does not depend on $e_i\in\mathcal{A}_i$. Indeed, if $g\gamma$ denote the image of the path γ under the action of $g\in G$, we have $|g\gamma|=|\gamma|$, $\phi(g\gamma)=\phi(\gamma)$. Summing for each $i=1,\dots,k$ over all edges in $\mathcal{A}_i$, we obtain

$$\begin{aligned}
A(\phi,e_i) &= \frac{1}{|\mathcal{A}_i|Q_i}\sum_{e\in\mathcal{A}_i}\sum_{\substack{\gamma\in\Gamma:\\ \gamma\ni e}}|\gamma|\phi(\gamma)\\
&= \frac{1}{|\mathcal{A}_i|Q_i}\sum_{e\in\mathcal{A}_i}\sum_{x,y}\sum_{\substack{\gamma\in\mathcal{G}(x,y):\\ \gamma\ni e}}\frac{d(x,y)K'(x,y)\pi'(x)}{\#\mathcal{G}(x,y)}\\
&\le \frac{1}{|\mathcal{A}_i|Q_i}\sum_{x,y}N_i(x,y)d(x,y)K'(x,y)\pi'(x).
\end{aligned}$$

This proves the desired bound.

EXAMPLE 4.2.3: Let $\mathcal{X}$ be the set of all the n-sets of $\{0,1,\dots,2n-1\}$. On this set, consider two chains. The unknown chain of interest is the chain K of Example 3.2.8:

$$K(x,y)=\begin{cases}1/n & \text{if } \#(x\cap y)=n-2 \text{ and } 0\in x\oplus y\\ 0 & \text{otherwise}\end{cases}$$

This is a reversible chain with respect to the uniform distribution $\pi\equiv\binom{2n}{n}^{-1}$. Let $\mathcal{A}_K=\{e=(x,y):K(x,y)\neq 0\}$ be the obvious K-adapted edge-set.

The better known chain K' that will be used for comparison is a special case of the chain considered of Example 3.2.7:

$$K'(x,y)=\begin{cases}1/n^2 & \text{if } \#(x\cap y)=n-2\\ 0 & \text{otherwise}\end{cases}$$

The chain K' is studied in detail in [34] using Fourier analysis on the Gelfand pair $(S_{2n},S_n\times S_n)$. The eigenvalues are known to be the numbers

$$\frac{i(2n-i+1)}{n^2} \quad \text{with multiplicity} \quad \binom{2n}{i}-\binom{2n}{i-1},\ \ 0\le i\le n.$$

In particular, the spectral gap of K' is $\lambda'=2/n$. This chain is known as the Bernoulli-Laplace diffusion model.

As in Example 3.2.8, the symmetric group S_{2n-1} which fixes 0 acts on $\mathcal{X}$ and preserves both chains K, K'. There are two classes $\mathcal{A}_1,\mathcal{A}_2$ of K-edges for this action: those edges (x,y), $x\oplus y=2$, with $0\in x\oplus y$ and those with $0\notin x\oplus y$. Hence, we have $\mathcal{E}'\le A\mathcal{E}$ with

$$A=\frac{2}{n^2\binom{2n}{n}}\max_{i=1,2}\left\{\sum_{\substack{x,y\\ x\oplus y=2}}N_i(x,y)d_K(x,y)\right\}.$$

Now, if $x\oplus y=2$ then

$$d_K(x,y)=\begin{cases}1 & \text{if } 0\in x\oplus y\\ 2 & \text{if } 0\notin x\oplus y.\end{cases}$$

Moreover, in both cases, $N_i(x,y)=0$ or 1. This yields

$$A\le\frac{4}{n^2\binom{2n}{n}}\sum_{\substack{x,y\\ x\oplus y=2}}1=4.$$

Thus

$$\mathcal{E}'\le 4\mathcal{E}.$$

This shows that

$$\lambda\ge\frac{1}{2n}$$

improving slightly upon the bound obtained in Example 3.2.8.

In their paper [34], Diaconis and Shahshahani actually show that

$$\|h_t'^x - 1\|_2 \le be^{-c} \quad \text{for} \quad t = \frac{1}{4}n(2c + \log n).$$

Using the comparison inequality $\mathcal{E}' \le 4\mathcal{E}$ and Theorem 4.1.1(2) we deduce from Diaconis and Shahshahani result that

$$|\!|\!| h_t - 1 |\!|\!| \le be^{-c} \quad \text{for} \quad t = n(2c + \log n).$$

Furthermore, the group S_{2n-1} fixing 0 acts with two transitive classes on $\mathcal{X}$. A vertex x is in one class or the other depending on whether or not x contains 0. The two classes have the same cardinality. Since $\|h_t^x - 1\|_2$ depends only of x through its class, we have

$$|\!|\!| h_t - 1 |\!|\!|^2 = \frac{1}{2}\left(\|h_t^{x_1} - 1\|_2^2 + \|h_t^{x_2} - 1\|_2^2\right)$$

where $x_1 \ni 0$ and $x_2 \not\ni 0$ are fixed elements representing their class. Hence, we also have

$$\max_x \|h_t^x - 1\|_2 \le 2be^{-c} \quad \text{for} \quad t = n(2c + \log n).$$

This example illustrates well the strength of the idea of comparison which allows a transfer of information from one example to another.

Bibliography

[1] Aldous D. (1983) *Random walks on finite groups and rapidly mixing Markov chains.* In Séminaire de Probabilités, XVII, LNM 986, Springer, Berlin.

[2] Aldous D. (1987) *On the Markov-chain simulation method for uniform combinatorial simulation and simulated annealing.* Prob. Eng. Info. Sci., 1, 33-46.

[3] Aldous D. and Fill J. (1996) Preliminary version of a book on finite Markov chains available via homepage http://www.stat.berkeley.edu/users/aldous

[4] Aldous D. and Diaconis P. (1987) *Strong uniform times and finite random walks.* Adv. Appl. Math. 8, 69-97.

[5] Alon N. (1986) *Eigenvalues and expanders.* Combinatorica 6, p.83-96.

[6] Alon N. and Milman V. (1985) λ_1, *isoperimetric inequalities for graphs and superconcentrators.* J. Comb. Th. B, 38, 78-88.

[7] Bakry D. and Emery M. (1985) *Diffusions hypercontractive.* Séminaire de probabilité XIX, Springer LNM 1123, 179-206.

[8] Bakry D. (1994) *L'hypercontractivité et son utilisation en théorie des semigroups.* In Ecole d'été de Saint Flour 1992, Springer LNM 1581.

[9] Bakry D., Coulhon T., Ledoux M., Saloff-Coste L. (1995) *Sobolev inequalities in disguise.* Indian Univ. Math. J., 44, 1043-1074.

[10] Bonami A. (1970) *Étude des coefficients de Fourier des fonctions de $L^p(G)$.* Ann. Inst. Fourier, 20, 335-402.

[11] Carlen E., Kusuoka S. and Stroock D. (1987) *Upper bounds for symmetric Markov transition functions.* Ann. Inst. H. Poincaré, Prob. Stat. 23, 245-287.

[12] Cheeger J. (1970) *A lower bound for the smallest eigenvalue of the Laplacian.* Problems in Analysis, Synposium in Honor of S. Bochner. Princeton University Press. 195-199.

[13] Chung F. and Yau S-T. (1994) *A harnack inequality for homogeneous graphs and subgraphs.* Communication in Analysis and Geometry, 2, 627-640.

[14] Chung F. and Yau S-T. (1995) *Eigenvalues of graphs and Sobolev inequalities.* Combinatorics, Probability and Computing, 4, 11-25.

[15] Davies E.B. (1989) *Heat kernels and spectral theory.* Cambridge University Press.

[16] Deuschel J-D. and Stroock D. (1989) *Large deviations.* Academic Press, Boston.

[17] Diaconis P. (1986) *Group representations in probability and statistics.* IMS, Hayward.

[18] Diaconis P. (1996) *The cutoff phenomenon in finite Markov chains.* Proc. Natl. Acad. Sci. USA, 93, 1659-1664.

[19] Diaconis P. and Fill J. (1990) *Strong stationary times via a new form of duality.* Ann. Prob. 18, 1483-1522.

[20] Diaconis P., Graham R. and Morrison J. (1990) *Asymptotic analysis of a random walk on a hypercube with many dimensions.* Random Structures and Algorithms, 1, 51-72.

[21] Diaconis P. and Gangolli A. (1995) *Rectangular arrays with fixed margins.* In Discrete Probability and Algorithms, (Aldous et al, ed.) 15-41. The IMA volumes in Mathematics and its Applications, Vol. 72, Springer-Verlag.

[22] Diaconis P. and Holmes S. (1995) *Three Examples of Monte-Carlo Markov Chains: at the Interface between Statistical Computing, Computer Science and Statistical Mechanics.* In Discrete Probability and Algorithms, (Aldous et al, ed.) 43-56. The IMA volumes in Mathematics and its Applications, Vol. 72, Springer-Verlag.

[23] Diaconis P. and Saloff-Coste L. (1993) *Comparison theorems for reversible Markov chains.* Ann. Appl. Prob. 3, 696-730.

[24] Diaconis P. and Saloff-Coste L. (1993) *Comparison techniques for random walk on finite groups.* Ann. Prob. 21, 2131-2156.

[25] Diaconis P. and Saloff-Coste L. (1994) *Moderate growth and random walk on finite groups.* G.A.F.A., 4, 1-36.

[26] Diaconis P. and Saloff-Coste L. (1995) *An application of Harnack inequalities to random walk on nilpotent quotients.*, J. Fourier Anal. Appl., Kahane special issue, 187-207.

[27] Diaconis P. and Saloff-Coste L. (1995) *Random walks on finite groups: a survey of analytical techniques.* In Probability on groups and related structures XI, H. Heyer (ed), World Scientific.

[28] Diaconis P. and Saloff-Coste L. (1996) *Nash inequalities for finite Markov chains.*, J. Th. Prob. 9, 459-510.

[29] Diaconis P. and Saloff-Coste L. (1996) *Logarithmic Sobolev inequalities and finite Markov chains.* Ann. Appl. Prob. 6, 695-750.

[30] Diaconis P. and Saloff-Coste L. (1995) *Walks on generating sets of Abelian groups.* Prob. Th. Rel. Fields. 105, 393-421.

[31] Diaconis P. and Saloff-Coste L. (1995) *What do we know about the Metropolis algorithm.* J.C.S.S. To appear.

[32] Diaconis D. and Shahshahani M. (1981) *Generating a random permutation with random transpositions.* Z. Wahrsch. Verw. Geb., 57, 159-179.

[33] Diaconis P. and Sahshahani M. (1987) *The subgroup algorithm for generating uniform random variables.* Probl. in Engin. Info. Sci., 1, 15-32.

[34] Diaconis P. and Shahshahani M. (1987) *Time to reach statinarity in the Bernoulli–Laplace diffusion model.* SIAM Jour. Math. Anal., 18, 208-218.

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

[36] Dinwoodie I. H. (1995) *A probability inequality for the occupation measure of a reversible Markov chain.* Ann. Appl. Prob., 5, 37-43.

[37] Dinwoodie I. H. (1995) *Probability inequalities for the occupation measure of a Markov chain.*

[38] Dyer M., Frieze A., Kannan R., Kapoor A., Perkovic L., and Vazirani U. (1993) *A mildly exponential time algorithm for approximating the number of solutions to a multidimensional knapsack problem.* Combinatorics, Probability and Computing, 2, 271-284.

[39] Dyer M. and Frieze A. (1991) *Computing the volume of convex bodies: a case where randomness provably helps.* Probabilistic Combinatorics and its applications, Proceedings of the AMS Symposia in Applied Mathematics 44, 123-170.

[40] Feller W. (1968) *An introduction to probability theory and its applications.* Vol. I, third edition, John Wiley & Sons, New-York.

[41] Fill J. (1991) *Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with application to the exclusion process.* Ann. Appl. Probab., 1, 62-87.

[42] Flatto L., Odlyzko A. and Wales D. (1985) *Random shuffles and group representations.* Ann. Prob. 13, 151-178.

[43] Frieze A., Kannan R. and Polson N. (1994) *Sampling from log concave distributions.* Ann. Appl. Prob. 4, 812-837.

[44] Gillman D. (1993) *Hidden Markov chains: rates of convergences and the complexity of inference.* Ph.D. thesis, Massachusets Institute of Technology, Department of mathematics.

[45] Gluck D. (1996) *Random walk and character ratios on finite groups of Lie type.* Adv. Math.

[46] Gross L. (1976) *Logarithmic Sobolev inequalities.* Amer. J. Math. 97, 1061-1083.

[47] Gross L. (1993) *Logarithmic Sobolev inequalities and contractivity properties of semigroups.* In Lecture Notes in Math. 1563. Springer.

[48] Higuchi Y. and Yoshida N. (1995) *Analytic conditions and phase transition for Ising models.* Lecture notes in Japanese.

[49] Hildebrand M. (1992) *Genrating random elements in $SL_n(F_q)$ by random transvections.* J. Alg. Combinatorics, 1, 133-150.

[50] Holley R. and Stroock D. (1987) *Logarithmic Sobolev inequalities and stochastic Ising models.* J. Stat. Phys. 46, 1159-1194.

[51] Horn R. and Johnson Ch. (1985) *Matrix analysis.* Cambridge University Press.

[52] Horn R and Johnson Ch. (1991) *Topics in Matrix analysis.* Cambridge University Press.

[53] Jerrum M. and Sinclair A. (1993)*Polynomial time approximation algorithms for the Ising model,* SIAM Journal of Computing, 22, 1087-1116.

[54] Jerrum M. and Sinclair A. (1997) *The Markov chain Monte Carlo method: an approach to approximate counting and integration.* In Approximation algorithms for NP-hard problems, D.S. Hochbaum (Ed.), PWS Publishing, Boston.

[55] Kahale N. (1995) *A semidefinite bound for mixing rates of Markov chains.* To appear in Random Structures and Algorithms.

[56] Kannan R. (1994) *Markov chains and polynomial time algorithms.* Proceedings of the 35th IEEE Symposium on Foundations of Computer Science, Computer Society Press, 656-671.

[57] Kemeny J. and Snell L. (1960) *Finite Markov chains.* Van Nostrand company, Princeton.

[58] Lawler G. and Sokal A. (1988) *Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger inequality.* Trans. AMS, 309, 557-580.

[59] Lezaud P. (1996) *Chernoff-type bound for finite Markov chains.* Ph.D. Thesis in progress, Université Paul Sabatier, Toulouse.

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

[61] Mann B. (1996) *Berry-Essen central limit theorems for Markov chains.* Ph.D. Thesis, Harvard University, Department of Mathematics.

[62] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A. and Teller E. (1953) *Equations of state calculations by fast computing machines.* J. Chem. Phys. 21, 1087-1092.

[63] Miclo L. (1996) *Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies.* To appear in Séminaire de Probabilité XXXI. Lecture notes in Math. Springer.

[64] Nash J. (1958) *Continuity of solutions of parabolic and elliptic equations.* Amer. J. Math. 80, 931-954.

[65] Rothaus O. (1980) *Logarithmic Sobolev inequalities and the spectrum of Sturm-Liouville operators.* J. Funct. Anal., 39, 42-56.

[66] Rothaus O. (1981) *Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators.* J. Funct. Anal., 42 , 110-120.

[67] Rothaus O. (1981) *Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities.* J. Funct. Anal., 42 , 102-109.

[68] Rothaus O. (1985) *Analytic inequalities, Isoperimetric inequalities and logarithmic Sobolev inequalities.* J. Funct. Anal., 64, 296-313.

[69] Saloff-Coste L. (1996) *Simple examples of the use Nash inequalities for finite Markov chains.* In Semstat III, Current Trends in Stochastic Geometry and its Applications. W.S. Kendall, O.E. Barndorff-Nielsen and MC. van Lieshout, Eds. Chapman & Hall.

[70] Senata E. (1981) *Non negative matrices and Markov chains* (2nd ed.) Springer.

[71] Sinclair A. (1992) *Improved bounds for mixing rates of Markov chains and multicommodity flow.* Combinatorics, Probability and Computing, 1, 351-370.

[72] Sinclair A. (1993) *Algorithms for random generation and counting: a Markov chain approach.* Birkhäuser, Boston.

[73] Stein E. and Weiss G. (1971) *Introduction to Fourier analysis in Euclidean spaces.* Princeton Univ. Press, Princeton.

[74] Stong R. (1995) *Random walks on the groups of upper triangular matrices* Ann. Prob., 23, 1939-1949.

[75] Stong R. (1995) *Eigenvalues of the natural random walk on the Burnside group $B(3, n)$.* Ann. Prob., 23, 1950-1960.

[76] Stong R. (1995) *Eigenvalues of random walks on groups.* Ann. Prob., 23, 1961-1981.

[77] Swendsen R. H. and Wang J-S. (1987) *Nonuniversal critical dynamics in Monte-Carlo simulations,* Physical review letters, 58, 86-88.

[78] Varopoulos N. (1985) *Semigroupes d'opérateurs sur les espaces L^p.* C. R. Acad. Sc. Paris. 301, Série I, 865-868.

[79] Varopoulos N. (1985) *Théorie du potentiel sur les groupes et les variétés.* C. R. Acad. Sc. Paris. 302, Série I, 203-205.