Università degli Studi Roma Tre
Facoltà di Scienze M.F.N.
Corso di Laurea in Matematica

Tesi di Laurea Magistrale in Matematica

The Cut-off Phenomenon for Monte Carlo Markov Chains

Candidato
Michele Salvi

Relatore
Prof. Fabio Martinelli

Anno Accademico 2008-2009
Ottobre 2009

Classificazione AMS: 60J10, 82B20, 82B26, 82C20, 60K35, 82C22, 82C80.
Parole chiave: Markov chains, cut-off, mixing time, Ising model, Glauber dynamics.

"How should I behave?"
Makoto

Acknowledgements

First of all I would like to offer deep thanks to my supervisor, Professor Fabio Martinelli, for his helpfulness in spite of his numerous engagements. He also played a significant role in reawakening my interest for Probability five years ago. I'm grateful also to the whole department of Mathematics of Roma3 University, who managed to create a positive atmosphere and a nice working climate between students and professors. A very special thank goes to Antonella Baldi for her pragmatism and her irony: without her now I would be a sad economist.

Many thanks to the members of my family, for their support, and thanks to my cats for their anti-stress function.

This thesis is dedicated to all who studied with me at least once.

Introduction

Markov chains were first introduced in 1906 by the Russian mathematician Andrey Markov. Their importance in the field of Probability grew along the whole past century, from the development of stochastic processes in the '20s, through the axiomatization of probability in the '30s, and into the 80's when the modern approach to Markov chains arose.

Particularly in the last thirty years, Markov processes assumed a central role of interest, finding applications in a huge number of different scientific fields, from genetics to theoretical computer science, from statistical physics to economics, from cryptography to Social sciences, just to name a few. To give an idea of the way they affect modern society, it is sufficient to say that the underlying algorithm of the web search engine Google is based on a particular Markov chain.

One of their most important applications is in computer simulations, which have revolutionized applied mathematics by providing ways to deal with high-dimensional intractable computations. One fundamental problem is: given a probability distribution π, how can we sample a random object with this distribution? The Monte Carlo method is an answer to this question: we can start a Markov chain with π as stationary distribution and let it run for enough time. Then the distribution of the last step of the chain will be a good approximation of π.

The 'classical' theory of Markov chains tries to quantify this *enough*, estimating how long we have to wait in order to reach a good approximation of the requested measure. We call this time the *mixing time* of the chain. The next step is to consider families of chains instead of one single process. The challenge is to understand how the mixing time grows as, for example, the size of the problem becomes bigger and bigger.

In this context, David Aldous and Persi Diaconis, in their article of 1986 ([1]), introduced for the first time the notion of *cut-off* to describe the sharp transition in convergence to stationarity for Markov chains modelling the shuffling of a card deck.

Take a sequence of Markov chains on the state spaces $\Omega^{(n)}$ with transition matrices $P^{(n)}$. Suppose that each chain is reversible with respect to a probability measure $\pi^{(n)}$ on $\Omega^{(n)}$, that is, $\forall x, y \in \Omega^{(n)}$

$$\pi^{(n)}(x)P^{(n)}(x,y) = \pi^{(n)}(y)P^{(n)}(y,x).$$

Define $d^{(n)}(t)$ as the distance (for example measured in total variation) of the n-th chain from $\pi^{(n)}$ at time t in the case of the worst starting point:

$$d^{(n)}(t) := \max_{x\in\Omega^{(n)}} \|P^{(n)}(x,\cdot) - \pi^{(n)}(\cdot)\|_{TV}.$$

The ε-mixing time of the n-th chain is the first time when $d^{(n)}$ becomes smaller than ε:

$$t_{mix}^{(n)}(\varepsilon) := \min\{t : d^{(n)}(t) \leq \varepsilon\}.$$

We say that this sequence of Markov chains exhibits a cut-off if, for any $0 < \varepsilon < \frac{1}{2}$,

$$\lim_{n\to\infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} = 1.$$

This means that, when the size of the problem (for example, in the shuffling card process it is the number of the cards in the deck) is very big, the distance from π goes from almost 1 (the maximum) to almost 0 in an interval around the mixing time that is relatively very small with respect to the mixing time itself.

Knowing the existence of the cut-off and the order of the mixing time is very useful for Monte Carlo simulations. In fact, we know that in order to have a good approximation of the measure π we have to run the chain at least for $O(t_{mix})$ steps; but we also know that waiting much longer is quite useless.

In the years after the article by Aldous and Diaconis, this particular behaviour was observed for a wide class of natural examples. Nevertheless, it is in general quite hard to prove rigorously the existence of a cut-off. The aim of this work is to present an overview of some of the latest results in this field.

In the first chapter, the basic theory on finite discrete-time Markov chains is summarized, stressing the notion of convergence to a stationary distribution. Then some useful techniques are presented, such as coupling and spectral analysis of the chains.

The second chapter introduces the definition of cut-off and cut-off window for ergodic Markov chains. After giving some basic results on this subject, an example of a chain that exhibits a cut-off is discussed: the simple random walk on the n-dimensional hypercube.

Chapter 3 is completely dedicated to birth-and-death processes. The most significant part is the proof of a necessary and sufficient condition for cut-off in such chains, according to the article [10] by Ding, Lubetzky and Peres, of 2008.

Chapter 4 deals with Glauber dynamics for the Ising model on the complete graph. Following [16] and [9], it is shown that in the high temperature regime this process has a cut-off at $t = \frac{n \log n}{2(1-\beta)}$ with a window of size n, while for any temperatures lower or equal to critical there is no cut-off and the order of mixing time changes.

Finally, in the last chapter, we change the subject a bit. We consider the Glauber dynamics treated in Chapter 4 in the continuous-time case and make an attempt to bound the Log-Sobolev constant with an analytical method, based on conditioning the entropy on the magnetization of the system.

Chapter 1

Background material

1.1 Finite Markov chains

1.1.1 Definition

A **stochastic process** is a family $\{X_t\}_{t\in I}$ of random variables indexed by a parameter $t \geq 0$ which is usually to be thought as a time parameter. The set I can be either discrete or continuous. In this work, except for the last chapter, we will analyze only a particular kind of stochastic processes, namely the Markov chains.

A **Markov chain** with state space Ω is a sequence of random variables $(X_0, X_1, ...)$ taking their values in Ω with a simple property: at time t it "chooses" its following position taking into account only its present position, X_t, and forgetting all its past moves. More precisely, $\forall t \in \mathbb{N}$, $\forall x, y \in \Omega$ and for all the possible values of $x_0, x_1, x_2, ..., x_{t-1} \in \Omega$, we have that

$$\begin{aligned} \mathbb{P}\left(X_{t+1} = y | X_0 = x_0, ..., X_{t-1} = x_{t-1}, X_t = x\right) &= \mathbb{P}\left(X_{t+1} = y | X_t = x\right) \\ &= P\left(x, y\right), \qquad (1.1) \end{aligned}$$

where $P(x, \cdot)$ is a fixed probability distribution. Equation (1.1) is called the **Markov property**. If Ω has a finite number of elements, then the $|\Omega|$x$|\Omega|$ matrix P whose elements P_{ij} are the probabilities $P(x_i, x_j)$ is called the **transition matrix**. Of course, $\forall x \in \Omega$,

$$\sum_{y\in\Omega} P(x, y) = 1,$$

so we say that P is a *stochastic* matrix. We will often call the whole Markov chain just P.

Let's call μ_t the $|\Omega|$-row-vector that describes the distribution of the chain at time t,

$$\mu_t(\cdot) = \mathbb{P}\left(X_t = \cdot\right);$$

if we condition on all possible values of the chain at time $t-1$, we have:

$$\mu_t(y) = \sum_{x\in\Omega} \mathbb{P}\left(X_{t-1} = x\right) P(x,y) = \sum_{x\in\Omega} \mu_{t-1}(x)P(x,y).$$

Rewriting this in vector form gives, for all t,

$$\mu_t = \mu_{t-1}P$$

and iterating

$$\mu_t = \mu_0 P^t, \tag{1.2}$$

where μ_0 is an arbitrary starting distribution. In words, multiplying the present-distribution-row-vector by P on the right gives the distribution of the next step. Since we will often deal with chains with the same transition probabilities but different starting distributions, we will use the notations $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ for, respectively, the probabilities and the expectations given $\mu_0 = \mu$. Similarly, $\mathbb{P}_x$ and $\mathbb{E}_x$ will consider a δ distribution in x as starting distribution, that is, the case in which we start the chain surely from the state x.

So, for example, starting in x, the probability of arriving in the state y after t steps will be

$$\mathbb{P}_x\left(X_t = y\right) = (\delta_x P^t)(y),$$

and this is equivalent to the (x,y)-th entry of the matrix P^t.

Finally, if we consider a $|\Omega|$-column-vector f, function of the states of Ω, and apply to the left the transition matrix, we obtain the expected value of f after one step:

$$Pf(x) = \sum_{y\in\Omega} P(x,y)f(y) = \sum_{y\in\Omega} f(y)\mathbb{P}_x\left(X_1 = y\right) = \mathbb{E}_x\left[f(X_1)\right].$$

1.1.2 Irreducibility, aperiodicity and regularity

In this section we are going to introduce two important properties of some Markov chains: aperiodicity and irreducibility. Almost every chain studied in this paper has these two properties, or can be slightly modified to become a chain of this kind, as we will see in the following.

The first property is very simple to understand: it just means that, starting from any other point of Ω, it is possible to reach any other state after some steps. To say it precisely, a Markov chain with transition matrix P is called **irreducible** if, $\forall x, y \in \Omega$, there exists a t (possibly depending on x and y) such that $P^t(x,y) > 0$.

Let's move to the second property. Let $T(x) := \{t \geq 1 : P^t(x,x) > 0\}$ be the set of times when, starting in x, it is possible to go back to x itself. The greatest common divisor of $T(x)$ is the **period** of the state x. If all states of the chain have period 1, then the chain is said to be **aperiodic**, otherwise it is periodic.

Finally we say that a chain is **regular** if exists an integer s such that $P^s(x,y) > 0$ for all $x, y \in \Omega$. It means that at time s we could be everywhere, no matter the starting point. Let's demonstrate a little result:

Proposition 1.1. *If P is irreducible and aperiodic, than it is regular.*

Proof: It's not hard to demonstrate that if $S \subset \mathbb{N}^+$, a set of non-negative integers closed under addition, has greatest common divisor 1, than there is some integer k such that $\forall j \geq k$, $j \in S$. Of course, for $x \in \Omega$, the set $T(x)$ is closed under addition: if $s, t \in T(x)$, then $P^{t+s}(x,x) \geq P^t(x,x)P^s(x,x) > 0$, and hence $s + t \in T(x)$. Besides, because of aperiodicity, the gcd of $T(x)$ is 1. Therefore there exists a $t(x)$ such that if $t \geq t(x)$ than $t \in T(x)$. By irreducibility we know that $\forall y \in \Omega$ there exists $r = r(x,y)$ such that $P^r(x,y) > 0$. Therefore, if we take $t \geq t(x) + r$, we have

$$P^t(x,y) \geq P^{t-r}(x,x)P^r(x,y) > 0.$$

This is true for every x and y. So, taking $t \geq \max_{x\in\Omega}\left(t(x) + \max_{y\in\Omega} r(x,y)\right)$, we obtain that $P^t(x,y) > 0$ for all $x, y \in \Omega$. ■

As we said above the problem of periodicity can be easily avoided with a simple trick: at each step we allow the chain to stand still with probability $\frac{1}{2}$. The new transition matrix will be $\tilde{P} = \frac{1}{2}Id + \frac{1}{2}P$, where Id is the identity matrix and P the original transition matrix. This way, our new Markov chain, called the **lazy version** of the original chain, will be obviously aperiodic, since $\forall x \in \Omega$ we have $\tilde{P}(x,x) > 0$. On the other hand, it's easy to verify that it preserves most of the original properties.

1.1.3 Stationary distributions

We say that a probability measure π on Ω is a **stationary distribution** for a Markov chain with transition matrix P if

$$\pi = \pi P, \tag{1.3}$$

or, rewriting this element-wise,

$$\pi(x) = \sum_{y\in\Omega} \pi(y)P(y,x) \qquad \forall x \in \Omega. \tag{1.4}$$

This means that if we are distributed according to π and perform a step of the chain, we are still distributed the same way.

These distributions will play a fundamental role in our study of the Markov chains: under very common conditions we have that as the time passes the chain approaches the stationary distribution in some sense. This is shown more precisely by the following famous theorem.

Theorem 1.2 (Ergodic theorem). *Let $\{X_t\}$ be a regular Markov chain with transition matrix P. Then there exists a unique stationary distribution π for the chain. Moreover*

$$P^t(x,y) \xrightarrow{t\to\infty} \pi(y) \qquad \forall x,y \in \Omega. \tag{1.5}$$

In particular, for every starting distribution μ_0 and for every function $h : \Omega \to \mathbb{R}$, we have

$$\mathbb{E}_{\mu_0}\left[h(X_t)\right] \xrightarrow{t\to\infty} \pi(h). \tag{1.6}$$

Proof: We will demonstrate an even stronger property that straight implies (1.5): there exists a unique stationary distribution π that verifies "$\exists\delta > 0$ such that

$$\max_{x\in\Omega}|P^t(x,y) - \pi(y)| \le \frac{1}{\delta}e^{-\delta t} \qquad \forall t \in \mathbb{N},\ \forall y \in \Omega". \tag{1.7}$$

Define

$$\begin{cases} M_y(t) := \max\limits_{x\in\Omega} P^t(x,y) \\ m_y(t) := \min\limits_{x\in\Omega} P^t(x,y) \end{cases}.$$

Obviously $m_y(t) \le M_y(t)$. Besides, we have

$$M_y(t+1) = \max_{x\in\Omega}\sum_{z\in\Omega} P(x,z)P^t(z,y) \le \max_{x\in\Omega}\sum_{z\in\Omega} P(x,z)M_y(t) = M_y(t)$$

and similarly $m_y(t+1) \ge m_y(t)$.

Let

$$\begin{cases} M_y(\infty) := \lim\limits_{t\to\infty} M_y(t) \\ m_y(\infty) := \lim\limits_{t\to\infty} m_y(t) \end{cases}.$$

so that $M_y(t) \searrow M_y(\infty)$ and $m_y(t) \nearrow m_y(\infty)$.

From the hypothesis of regularity we know that $\exists t_0 \in \mathbb{N}$ and $\varepsilon > 0$:

$$P^{t_0}(x,y) \ge \varepsilon \qquad \forall x,y \in \Omega.$$

So, taking $t \geq t_0$ we have

$$\begin{aligned} P^t(x,y) &= \sum_{z\in\Omega} P^{t_0}(x,z)P^{t-t_0}(z,y) \\ &= \sum_{z\in\Omega} \left[P^{t_0}(x,z) - \varepsilon P^{t_0}(y,z)\right] P^{t-t_0}(z,y) + \varepsilon P^t(y,y) \\ &\leq M_y(t-t_0) \sum_{z\in\Omega} \left[P^{t_0}(x,z) - \varepsilon P^{t_0}(y,z)\right] + \varepsilon P^t(y,y) \\ &= (1-\varepsilon)M_y(t-t_0) + \varepsilon P^t(y,y). \end{aligned}$$

Since this is true for every $x \in \Omega$, we have also

$$M_y(t) \leq (1-\varepsilon)M_y(t-t_0) + \varepsilon P^t(y,y).$$

The very same way one can see that

$$m_y(t) \geq (1-\varepsilon)M_y(t-t_0) + \varepsilon P^t(y,y).$$

Putting together these two inequalities we obtain

$$0 \leq M_y(t) - m_y(t) \leq (1-\varepsilon)[M_y(t-t_0) - m_y(t-t_0)];$$

iterating this, if $t \geq kt_0,\ k \in \mathbb{N}$,

$$M_y(t) - m_y(t) \leq (1-\varepsilon)^k[M_y(t-kt_0) - m_y(t-kt_0)].$$

Write $t = \lfloor \frac{t}{t_0} \rfloor t_0 + l,\ \ 0 \leq l \leq t_0$; then

$$M_y(t) - m_y(t) \leq (1-\varepsilon)^{\lfloor \frac{t}{t_0} \rfloor}[M_y(l) - m_y(l)] \leq (1-\varepsilon)^{\lfloor \frac{t}{t_0} \rfloor} \leq \frac{1}{\delta}e^{-\delta n}$$

for some $\delta > 0$. Now the vector π with $\pi(y) = \lim_{t\to\infty} M_y(t) = \lim_{t\to\infty} m_y(t)$ is well defined. It remains to show that π is actually a stationary distribution for P and that it is the only one.

Since, $\forall t > 0$ and $x \in \Omega$, $\sum_y P^t(x,y) = 1$, then $\sum_y \pi(y) = 1$, so π is a probability vector. It is also invariant:

$$\sum_{y\in\Omega} \pi(y)P(y,x) = \lim_{t\to\infty} \sum_{y\in\Omega} P^t(z,y)P(y,x) = \lim_{t\to\infty} P^{t+1}(z,x) = \pi(x).$$

If we have another probability vector μ such that $\mu P = \mu$, then, $\forall t > 0$, $\mu P^t = \mu$ and

$$\mu(x) = \lim_{t\to\infty} \sum_y \mu(y)P^t(y,x) = \sum_y \mu(y)\pi(x) = \pi(x),$$

so that $\mu = \pi$. Finally, since we know that for every possible probability vector μ_0 we have that $\mu_0 P^t \to \pi$, for every function h

$$\mathbb{E}_{\mu_0}\left[h(X_t)\right] = \sum_x \sum_y \mu_0(y)P^t(y,x)h(x) \xrightarrow{t\to\infty} \sum_x \pi(x)h(x) = \pi(h).$$

■

Because of this theorem, we can also call π the **equilibrium distribution**.

1.1.4 Reversibility

We say that a distribution of probability π satisfies the **detailed balance equations** for a Markov chain P if, $\forall x, y \in \Omega$,

$$\pi(x)P(x,y) = \pi(y)P(y,x). \tag{1.8}$$

This property is often useful to verify that a probability distribution is stationary:

Proposition 1.3. *If π verifies the detailed balance equations for a chain P, then it is stationary for that chain.*

Proof: We have just to sum over all $y \in \Omega$ on both sides of (1.8) and remember that P is a stochastic matrix. ■

Iterating (1.8) we obtain, for any sequence $x_0, ...x_t \in \Omega$,

$$\pi(x_0)P(x_0,x_1)...P(x_{t-1},x_t) = \pi(x_n)P(x_t,x_{t-1})...P(x_1,x_0).$$

This means that, starting from the stationary measure, the distribution of the variables $(X_0, ..., X_t)$ and that of $(X_t, ..., X_0)$ is the same. This is why such kind of Markov chains are called **reversible**.

1.2 Convergence to equilibrium

1.2.1 Total variation distance

We have shown with the Ergodic Theorem (Theorem 1.2) that Markov chains, under some conditions, in the infinite time limit "look like" their own stationary distributions. A natural question that arises is: how *fast* is this convergence? After how many steps we can consider the chain *close* to its equilibrium?

To answer we need first of all to define what the word "close" mean, that is, we have to define a metric in the contest of probability measures.

There are many choices we can do to define a distance on the space of all probability measures on the discrete space Ω. Widely used examples of such objects are the **l^p-distances**, useful especially for an analytic approach to the study of Markov chains. Given an integer $p \geq 1$ and two probability measures μ (thought as the 'reference measure') and ν on Ω, the l^p distance deals with the density of ν respect to μ:

$$\left\|\frac{\nu}{\mu} - 1\right\|_{l^p} = \left(\sum_{x\in\Omega}\left|\frac{\nu(x)}{\mu(x)} - 1\right|^p\right)^{\frac{1}{p}}$$

and

$$\left\| \frac{\nu}{\mu} - 1 \right\|_{l^\infty} = \max_{x \in \Omega} \left| \frac{\nu(x)}{\mu(x)} - 1 \right|.$$

Another natural metric we can define on the same space is the one inducted by the **total variation distance**:

$$\|\mu - \nu\|_{TV} := \max_{A \subset \Omega} |\mu(A) - \nu(A)| = \max_{A \subset \Omega} \left| \sum_{x \in A} (\mu(x) - \nu(x)) \right|. \tag{1.9}$$

The interpretation is probabilistic: take the event on which the difference between μ and ν is maximum; then this difference is exactly the TVd. The next proposition gives a useful characterization:

Proposition 1.4. *Let μ and ν be two probability distributions on Ω. Then the total variation distance between them is exactly one half the l^1 distance, that is*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \tag{1.10}$$

Proof: Let $B := \{x \in \Omega : \mu(x) \geq \nu(x)\}$. For any event A we have

$$\mu(A) - \nu(A) = \mu(A \cap B) - \nu(A \cap B) + \mu(A \cap B^c) - \nu(A \cap B^c) \leq \mu(B) - \nu(B)$$

since, by definition of B, $\mu(A \cap B^c) - \nu(A \cap B^c)$ is negative and in the second inequality we have just added a positive term $(\mu(A^c \cap B) - \nu(A^c \cap B))$. Analogously, for any event A',

$$\nu(A') - \mu(A') \leq \nu(B^c) - \mu(B^c).$$

Note that

$$(\mu(B) - \nu(B)) - (\nu(B^c) - \mu(B^c)) = \mu(\Omega) - \nu(\Omega) = 0$$

so that the two upper bounds we have found have the same value. Furthermore, if we take $A = B$ and $A' = B^c$, then $|\mu(A) - \nu(A)|$ is exactly equal to the upper bounds. Thus

$$\|\mu - \nu\|_{TV} = \frac{1}{2}(\mu(B) - \nu(B)) + \frac{1}{2}(\nu(B^c) - \mu(B^c)) = \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

■

Note that from this characterization follows immediately the triangle inequality:

$$\begin{aligned} \|\mu - \nu\|_{TV} &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| \leq \frac{1}{2} \sum_{x \in \Omega} (|\mu(x) - \eta(x)| + |\eta(x) - \nu(x)|) \\ &= \|\mu - \eta\|_{TV} + \|\eta - \nu\|_{TV} \end{aligned}$$

for any probability measures μ, ν and η on Ω. If we had any doubt, now we are sure that the total variation is actually a distance. Finally we give a further way of describing the TVd between two measures:

Proposition 1.5. *Let μ and ν be two probability distributions on Ω. Then*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup_{f:\|f\|_\infty \leq 1} |\mu(f) - \nu(f)| . \tag{1.11}$$

Proof: $\forall f$ such that $\|f\|_\infty := \max_{x\in\Omega} |f(x)| \leq 1$,

$$\begin{aligned} \frac{1}{2}\left|\sum_{x\in\Omega} f(x)\mu(x) - \sum_{x\in\Omega} f(x)\nu(x)\right| &\leq \frac{1}{2}\sum_{x\in\Omega} |f(x)(\mu(x) - \nu(x))| \\ &\leq \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)| \\ &\leq \|\mu - \nu\|_{TV}, \end{aligned} \tag{1.12}$$

which shows the $(\geq)$ side. For the reverse, set

$$f^*(x) := \begin{cases} +1 \text{ if } \mu(x) \geq \nu(x) \\ -1 \text{ if } \mu(x) < \nu(x). \end{cases}$$

Then

$$\begin{aligned} \left|\sum_{x\in\Omega} f^*(x)(\mu(x) - \nu(x))\right| &= \sum_{\mu(x)\geq\nu(x)} (\mu(x) - \nu(x)) + \sum_{\mu(x)<\nu(x)} (\nu(x) - \mu(x)) \\ &= \sum_{x\in\Omega} |\mu(x) - \nu(x)|, \end{aligned} \tag{1.13}$$

and Lemma 1.4 ends the proof. ■

1.2.2 Distance from equilibrium

From now on, we will call $d(t)$ the distance of a Markov chain P at time t from its stationary distribution π starting from the worst initial distribution:

$$d(t) := \max_{\mu} \|\mathbb{P}_\mu (X_t = \cdot) - \pi(\cdot)\|_{TV}$$

Note that

$$\begin{aligned} \max_{\mu} \frac{1}{2}\sum_{y\in\Omega} |\mathbb{P}_\mu (X_t = y) - \pi(y)| &= \max_{\mu} \frac{1}{2}\sum_{y\in\Omega}\left|\sum_{x\in\Omega}(\mu(x)P^t(x,y) - \mu(x)\pi(y))\right| \\ &\leq \max_{\mu} \frac{1}{2}\sum_{x\in\Omega}\sum_{y\in\Omega} |P^t(x,y) - \pi(y)|\mu(x) \end{aligned}$$

$$\leq \max_{\mu} \frac{1}{2} \max_{x\in\Omega} \left[\sum_{y\in\Omega} |P^t(x,y) - \pi(y)| \right] \sum_{x\in\Omega} \mu(x)$$

$$= \max_{x\in\Omega} \left[\frac{1}{2} \sum_{y\in\Omega} |P^t(x,y) - \pi(y)| \right] = \max_{x\in\Omega} \|P^t(x,\cdot) - \pi(\cdot)\|_{TV}.$$

Since the maximum over all possible distribution is made also on the point masses, the inverse inequality holds too. It follows that we can take just the worst starting point instead of the worst starting distribution:

$$d(t) := \max_{x\in\Omega} \|P^t(x,\cdot) - \pi(\cdot)\|_{TV}. \tag{1.14}$$

As one could expect the function $d(t)$ is decreasing in time, as follows from next proposition.

Proposition 1.6. *Let P be the transition matrix of a Markov chain with state space Ω and let μ, ν be any two probability distributions on Ω. Then*

$$\|\mu P - \nu P\|_{TV} \leq \|\mu - \nu\|_{TV} \tag{1.15}$$

Proof:

$$\frac{1}{2} \sum_{x\in\Omega} |\mu P(x) - \nu P(x)| = \frac{1}{2} \sum_{x\in\Omega} \left| \sum_{y\in\Omega} \mu(y)P(y,x) - \nu(y)P(y,x) \right|$$
$$\leq \frac{1}{2} \sum_{y\in\Omega} \sum_{x\in\Omega} P(y,x)|\mu(y) - \nu(y)| = \|\mu - \nu\|_{TV}$$

■

Corollary 1.7. *Let P be the transition matrix of a Markov chain with stationary distribution π. Then, for any $t \geq 0$*

$$d(t+1) \leq d(t). \tag{1.16}$$

Proof: For any $x \in \Omega$, taking in Proposition 1.6 $\mu := \delta_x P^t$ and $\nu := \pi P^t = \pi$, we have

$$\|P^{t+1}(x,\cdot) - \pi(\cdot)\|_{TV} \leq \|P^t(x,\cdot) - \pi(\cdot)\|_{TV}.$$

So, if $\bar{x}$ is a state that realizes the maximum $\max_{x\in\Omega} \|P^{t+1}(x,\cdot) - \pi(\cdot)\|_{TV}$, we have

$$d(t+1) = \max_{x\in\Omega} \|P^{t+1}(x,\cdot) - \pi(\cdot)\|_{TV} = \|P^{t+1}(\bar{x},\cdot) - \pi(\cdot)\|_{TV}$$

$$\le \|P^t(\bar{x},\cdot) - \pi(\cdot)\|_{TV} \le \max_{x\in\Omega} \|P^t(x,\cdot) - \pi(\cdot)\|_{TV}$$
$$= d(t).$$

■

In section 1.1.3 we have seen that every irreducible and aperiodic (regular would be sufficient) Markov chain has its stationary measure and that the distribution of the process is exactly this measure in the limit for the time going to infinity. In the proof of the Ergodic Theorem 1.2, also known as the Perron-Frobenius theorem, it was even shown that this convergence has an exponential rate. As a direct consequence of this fact, we have that also the total variation distance of the chain from the stationary distribution has this behaviour:

Corollary 1.8 (of Theorem 1.2). *Let P be an irreducible and aperiodic Markov chain with stationary distribution π. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that*

$$\max_{x\in\Omega} \|P^t(x,\cdot) - \pi(\cdot)\|_{TV} \le C\alpha^t. \tag{1.17}$$

Proof:

$$\begin{aligned}\max_{x\in\Omega} \|P^t(x,\cdot) - \pi(\cdot)\|_{TV} &= \max_{x\in\Omega} \frac{1}{2}\sum_{y\in\Omega} |P^t(x,y) - \pi(y)| \\ &\le \max_{x\in\Omega} \frac{1}{2}\sum_{y\in\Omega} \max_{x,y\in\Omega} \left(|P^t(x,y) - \pi(y)|\right) \\ &= \frac{n}{2} \max_{x,y\in\Omega} \left(|P^t(x,y) - \pi(y)|\right),\end{aligned}$$

where $n := |\Omega|$. From equation (1.7) we know that there exists a $\delta > 0$ such that for any $y \in \Omega$, $t \ge 0$,

$$\max_{x,y\in\Omega} \left(|P^t(x,y) - \pi(y)|\right) \le \frac{1}{\delta} e^{-\delta t},$$

so that setting $C := \frac{n}{2\delta}$ and $\alpha := e^{-\delta}$ we have (1.17). ■

Theorem 1.2 and Corollary 1.8 ensure that one day our Markov chain will be very close to stationarity. But since we don't know anything about the constants $(\delta, C, \alpha ...)$ involved in the statements, we can not say *how much* we have to wait to be sufficiently close.

A good branch of the studies on Markov chains in the last decades concentrated its efforts in understanding more in detail this kind of problems.

1.2.3 The mixing time

Suppose we want to know how long we have to wait before some Markov chain P reaches a distance from its stationary distribution of ϵ. It would be useful to have a time-parameter that formalizes this concept. We define the ε-mixing time of the chain as

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}. \tag{1.18}$$

A chain is said rapidly mixing if $t_{mix}(\varepsilon)$ is polynomial in $\log\left(\frac{1}{\varepsilon}\right)$ and the size of the problem.

By convention if we take $\varepsilon = \frac{1}{4}$, we call it simply **mixing time**:

$$t_{mix} := t_{mix}\left(\frac{1}{4}\right) = \min\left\{t : d(t) \leq \frac{1}{4}\right\}. \tag{1.19}$$

The study of the only t_{mix} is often sufficient, because of the next result:

Proposition 1.9. *Let P be the transition matrix of an irreducible, aperiodic Markov chain with $t_{mix}(c) \leq T$ for some $c < \frac{1}{2}$. Then, for this Markov chain,*

$$t_{mix}(\varepsilon) \leq \left\lceil \frac{\log \varepsilon}{\log(2c)} \right\rceil T. \tag{1.20}$$

In particular

$$t_{mix}(\varepsilon) \leq \left\lceil \log_2 \varepsilon^{-1} \right\rceil t_{mix}. \tag{1.21}$$

The proof of this fact requires the notion of coupling, thus it is put off to Section 1.3.4.

1.3 Some techniques

1.3.1 The specrtal gap and the Dirichlet form

An important and powerful method to analyze the convergence to equilibrium of a Markov chain is the study of the eigenvalues of its transition matrix. First of all let's see some basic properties of these eigenvalues.

Proposition 1.10. *Let P be the transition matrix of a finite Markov chain. If λ is an eigenvalue of P, then $|\lambda| \leq 1$.*

Proof: For any function $f : \Omega \to \mathbb{R}$, the infinity-norm is defined as $\|f\|_\infty := \max_{x\in\Omega} |f(x)|$. For any $x \in \Omega$ we have

$$|Pf(x)| = \Big|\sum_{y\in\Omega} P(x,y)f(y)\Big| \leq \|f\|_\infty \Big|\sum_{y\in\Omega} P(x,y)\Big| = \|f\|_\infty.$$

Taking the $\max_{x \in \Omega}$ we have

$$\|Pf\|_\infty \leq \|f\|_\infty.$$

Thus, if ϕ is the eigenfunction corresponding to the eigenvalue λ, choosing $f = \phi$,

$$\|\phi\|_\infty \geq \|P\phi\|_\infty = \|\lambda\phi\|_\infty = \lambda\|\phi\|_\infty.$$

■

It is also possible to prove that if P is irreducible, then the vector $\bar{1} := (1, 1, ..., 1)$ generates the vector space of the eigenfunctions corresponding to the eigenvalue 1, and that if P is also aperiodic, then -1 is not an eigenvalue of P.

Proposition 1.11. *If $\tilde{P}$ is the transition matrix of the lazy version of a chain with transition matrix P, then all the eigenvalues of $\tilde{P}$ are non-negative.*

Proof: Let f be an eigenfunction of $\tilde{P} = \frac{P+Id}{2}$ with eigenvalue λ. Then

$$\lambda f = \tilde{P} f = \frac{Pf + f}{2}.$$

It follows that $(2\lambda - 1)$ is an eigenvalue of P. Thus $2\lambda - 1 \geq -1$, and so $\lambda \geq 0$. ■

For a reversible transition matrix P we label the eigenvalues in a decreasing order:

$$1 = \lambda_1 > \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1. \tag{1.22}$$

Define

$$\lambda_* := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P,\ \lambda \neq 1\}. \tag{1.23}$$

We call the difference $\gamma_* := 1 - \lambda_*$ the **absolute spectral gap**, while the difference $\gamma := 1 - \lambda_2$ is just the **spectral gap**. If P is aperiodic and irreducible, then $\gamma_* > 0$ since, as we said, there is no -1-eigenvalue. Of course, by Proposition 1.11, whenever P is a lazy chain, $\gamma_* = \gamma$.

Define now on the space $\mathbb{R}^\Omega$ the inner product

$$\langle f, g \rangle_\pi := \sum_{x \in \Omega} f(x) g(x) \pi(x). \tag{1.24}$$

We will write $f \perp_\pi g$ if $\langle f, g \rangle_\pi = 0$. A well known property of reversible chains is that the inner product space $(\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\pi)$ has an orthonormal basis of real-valued eigenfunctions $\{f_j\}_{J=1}^n$ corresponding to the real eigenvalues $\{\lambda_j\}$. It follows that any function $f : \Omega \to \mathbb{R}$ can be decomposed as

$$P^t f = \sum_{j=1}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t. \tag{1.25}$$

In this case we can define the **Dirichlet form** associated to our chain as

$$\mathcal{D}(f,g) := \langle (Id-P)f, g\rangle_\pi$$

where f and g are functions on Ω. In particular, we define $\mathcal{D}(f) := \mathcal{D}(f,f)$. Just manipulating the definition and using the reversibility we can rewrite it as

$$\mathcal{D}(f) = \frac{1}{2}\sum_{x,y\in\Omega}[f(x)-f(y)]^2\pi(x)P(x,y). \tag{1.26}$$

We introduced the Dirichlet form for the following useful characterization of the spectral gap:

Lemma 1.12.

$$\gamma = \min_{\substack{f\in\mathbb{R}^\Omega \\ f\perp_\pi \bar{1}, \|f\|_2=1}} \{\mathcal{D}(f)\} \tag{1.27}$$

Proof: Say $|\Omega| = n$. Let $\{f_1, ..., f_n\}$ be an orthonormal basis of eigenfunctions for $(\mathbb{R}^\Omega, \langle\cdot,\cdot\rangle_\pi)$, with f_k associated to the eigenvalue λ_k for all k. Take $f_1 = \bar{1}$. By (1.25) we know that any function f can be written as $\sum_j \langle f, f_j\rangle_\pi f_j$ and then the l^2 norm of f is

$$\|f\|_2^2 = \langle f, f\rangle_\pi = \sum_{j=1}^n |\langle f, f_j\rangle_\pi|^2.$$

Whenever $\|f\|_2 = 1$ and $f \perp_\pi \bar{1}$ we can write f as $f = \sum_{j=2}^n a_j f_j$ with $\sum_{j=2}^n a_j^2 = 1$. Thus

$$\langle (Id-P)f, f\rangle_\pi = \sum_{j=2}^n a_j^2(1-\lambda_j) \geq (1-\lambda_2).$$

Finally take $f = f_2$ to realize the minimum. ■

Notice that equivalent expressions are

$$\gamma = \min_{\substack{f\in\mathbb{R}^\Omega \\ f\perp_\pi \bar{1}, f\neq 0}} \left\{\frac{\mathcal{D}(f)}{\|f\|_2^2}\right\}, \tag{1.28}$$

since $\tilde{f} = \frac{f}{\|f\|_2}$ has l^2 norm equal to one and satisfies $\mathcal{D}(\tilde{f}) = \frac{\mathcal{D}(f)}{\|f\|_2^2}$, and

$$\gamma = \min_{\substack{f\in\mathbb{R}^\Omega \\ Var_\pi(f)\neq 0}} \left\{\frac{\mathcal{D}(f)}{Var_\pi(f)}\right\} \tag{1.29}$$

(where $Var_\pi(f) = \sum f^2(x)\pi(x) - (\sum f(x)\pi(x))^2$), just replacing the f of the lemma with $\tilde{f} = f - \mathbb{E}_\pi[f]$.

1.3.2 The relaxation time

How can we relate the eigenvalues of the transition matrix to our study of the speed of convergence of the chain?
The **relaxation time** of a reversible Markov chain is defined as

$$t_{rel} := \frac{1}{\gamma_*}. \tag{1.30}$$

One possible link between t_{mix} and t_{rel} is the inequality

$$Var_\pi(P^t f) \leq (1-\gamma_*)^{2t} Var_\pi(f), \tag{1.31}$$

that we don't prove here. But the most interesting relation is shown in the next theorem.

Theorem 1.13. *Let P be the transition matrix of a reversible, irreducible and aperiodic Markov chain with state space Ω and let $\pi_{min} := \min_{x\in\Omega} \pi(x)$. Then*

$$(t_{rel}-1)\log\left(\frac{1}{2\varepsilon}\right) \leq t_{mix}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon\pi_{min}}\right) t_{rel}. \tag{1.32}$$

Proof: For the first inequality suppose that f is an eigenfunction of P with eigenvalue $\lambda \neq 1$. Since the eigenfunctions are orthogonal with respect to the inner product $\langle\cdot,\cdot\rangle_\pi$ and since $\bar{1}$ is an eigenfunction, it follows $\sum_y \pi(y)f(y) = \langle\bar{1}, f\rangle_\pi = 0$. Therefore

$$|\lambda^t f(x)| = |P^t f(x)| = \Big|\sum_{y\in\Omega}[P^t(x,y)f(y) - \pi(y)f(y)]\Big| \leq \|f\|_\infty 2d(t).$$

Taking the state x which realizes the infinity norm and setting $t = t_{mix}(\varepsilon)$ gives

$$|\lambda^{t_{mix}(\varepsilon)}| \leq 2d(t_{mix}(\varepsilon)) \leq 2\varepsilon.$$

Whence

$$t_{mix}(\varepsilon)\left(\frac{1}{|\lambda|}-1\right) \geq t_{mix}(\varepsilon)\log\left(\frac{1}{|\lambda|}\right) \geq \log\left(\frac{1}{2\varepsilon}\right).$$

Minimizing the left hand side over the eigenvalues different from 1 and rearranging gives the bound.

Now the second (and more useful) inequality. By Proposition 1.5 we can write

$$\|P^t(x,\cdot)-\pi\|_{TV} = \frac{1}{2}\sup_{g:|g|\leq 1}|P^t_x(g)-\pi(g)| = \frac{1}{2}\sup_{g:|g|\leq 1}|P^t_x(\tilde{g})| \tag{1.33}$$

where $\tilde{g}$ is the function g minus its mean. For all functions $\tilde{g}$ with 0-mean we have

$$\begin{aligned} |P_x^t(\tilde{g})| &\leq \frac{1}{\pi(x)} \sum_{y \in \Omega} \pi(y) |P_y^t(g)| \\ &\leq \frac{1}{\pi(x)} Var_\pi(P^t(\tilde{g}))^{\frac{1}{2}} \\ &\leq \frac{1}{\pi(x)} e^{-t\gamma^*} Var_\pi(g)^{\frac{1}{2}} \end{aligned} \tag{1.34}$$

where we have used Schwartz inequality and equation (1.31). Remember that $\|g\|_\infty \leq 1$ and hence $Var(g) \leq \pi(g^2) \leq 1$. Summarizing

$$\max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \leq \frac{1}{\pi_*} e^{-t\gamma^*}.$$

The right hand side is smaller than ε if

$$t \geq \frac{1}{\gamma^*} \log\left(\frac{1}{\varepsilon \pi_*}\right),$$

so that t_{mix} has to be smaller than the required quantity. ■

Most of the times finding precisely the eigenvalues of a transition matrix on a big state space turns out to be almost impossible. Some techniques to bound the spectral gap of a chain will be shown in the next chapters.

1.3.3 Coupling

A **coupling** of two probability distribution μ and ν on a state space Ω is a pair of random variables (X, Y), defined on $\Omega \times \Omega$, such that their marginal distributions are respectively μ and ν:

$$\begin{aligned} \sum_{y \in \Omega} \mathbb{P}(X = x, Y = y) &= \mu(x), \\ \sum_{x \in \Omega} \mathbb{P}(X = x, Y = y) &= \nu(y). \end{aligned}$$

There always exists a coupling between any two probability distribution, for example letting X and Y be independent. However very interesting properties come out when we force the two random variables to assume the same values. Next proposition, for example, gives a nice and useful characterization of the total variation distance based on the concept of coupling.

Proposition 1.14. *Given two probability distributions μ and ν on Ω, we have that*

$$\|\mu - \nu\|_{TV} = \inf \{\mathbb{P}(X \neq Y) : (X, Y) \textit{ is a coupling of } \mu \textit{ and } \nu\}. \tag{1.35}$$

Proof: First note that for every coupling (X,Y) of μ and ν and for any event $A \subseteq \Omega$

$$\begin{aligned}\mu(A)-\nu(A) &= \mathbb{P}(X\in A)-\mathbb{P}(Y\in A)\\ &= \mathbb{P}(X\in A, Y\notin A)+\mathbb{P}(X\in A, Y\in A)-\mathbb{P}(Y\in A)\\ &\le \mathbb{P}(X\in A, Y\notin A)\\ &\le \mathbb{P}(X\neq Y).\end{aligned}$$

Taking the maximum over all possible events on the left side and the infimum over all possible coupling on the right we obtain:

$$\|\mu-\nu\|_{TV} \le \inf\{\mathbb{P}(X\neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\}. \qquad (1.36)$$

Now we'll construct a coupling such that equality holds, forcing X and Y to be equal as often as possible. To do that, $\forall x \in \Omega$ we will assign value x to both X and Y with the highest probability allowed, that is $\mu(x)\wedge\nu(x)$.

Set

$$p=\sum_{x\in\Omega}\mu(x)\wedge\nu(x)=\sum_{\substack{x\in\Omega:\\ \mu(x)\le\nu(x)}}\mu(x)+\sum_{\substack{x\in\Omega:\\ \nu(x)<\mu(x)}}\nu(x).$$

Adding and subtracting $\sum_{x:\mu(x)>\nu(x)}\mu(x)$ to the right-hand side gives

$$p=1-\sum_{\substack{x\in\Omega:\\ \nu(x)<\mu(x)}}(\mu(x)-\nu(x))=1-\|\mu-\nu\|_{TV},$$

where the last equality follows immediately from the proof of Proposition 1.4.

Now flip a coin with probability of heads equal to p. If the coin comes up heads then we will take $X=Y=x$ with probability

$$\frac{\mu(x)\wedge\nu(x)}{p};$$

if the coin comes up tails, we will choose X according to the probability distribution

$$d_X(x)=\begin{cases}\dfrac{\mu(x)-\nu(x)}{1-p} & \text{if } \mu(x)>\nu(x)\\ 0 & \text{otherwise}\end{cases}$$

and independently Y according to the probability distribution

$$d_Y(x)=\begin{cases}\dfrac{\nu(x)-\mu(x)}{1-p} & \text{if } \nu(x)>\mu(x)\\ 0 & \text{otherwise.}\end{cases}$$

With simple calculations it is possible to verify that d_X and d_Y are actually probability distributions and that with this choice the marginals of X and Y are μ and ν. Finally note that $X = Y$ if and only if the coin shows tails, so that

$$\mathbb{P}(X \neq Y) = 1 - p = \|\mu - \nu\|_{TV}.$$

■

1.3.4 Coupling of Markov chains

Let's take two Markov chains defined on the same state space Ω, both with transition matrix P, but with different starting positions chosen according to μ_0 and ν_0. We would like to couple, in the sense of the last section, the distributions of the two chains at each step.

A coupling of these two chains is a process $(X_t, Y_t)_{t\geq 0}$ such that $X_t \sim P^t_{\mu_0}$ and $Y_t \sim P^t_{\nu_0}$. If X_t and Y_t start from the points x and y, we will use the notation $\mathbb{P}_{x,y}$ for the probability on the space where both the chains are defined.

Again we will be interested in forcing the two processes to be in the same point of Ω. Besides, once they have met, we will keep them together forever:

$$X_s = Y_s \Rightarrow X_t = Y_t \quad \forall t \geq s. \tag{1.37}$$

This can be easily achieved choosing next position for X_t according to P and then making the same choice for Y_t.

A very useful result provided by the coupling method is a bound for the distance from the stationarity, as stated by the Corollary of the following theorem.

Theorem 1.15. *Given a transition matrix P on the space Ω, consider two Markov chains starting from two different states x, y and evolving according to P. Take any coupling (X_t, Y_t) of the chains verifying (1.37), and define*

$$\tau_c := \min\{t : X_t = Y_t\},$$

the first time the two copies meet. Then

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \mathbb{P}_{x,y}(\tau_c > t). \tag{1.38}$$

Proof: Since at each step (X_t, Y_t) is a coupling of the measures $P^t(x,\cdot)$ and $P^t(y,\cdot)$ by construction, Proposition 1.14 gives immediately

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \mathbb{P}_{x,y}(X_t \neq Y_t).$$

Furthermore (1.37) implies $\mathbb{P}_{x,y}(X_t \neq Y_t) = \mathbb{P}_{x,y}(\tau_c > t)$, so we are done. ■

Corollary 1.16. *Suppose that for each pair of state $x, y \in \Omega$ there is a coupling $(X_t, Y_t)_{t\geq 0}$ such that $X_0 = x, Y_0 = y$. Then*

$$d(t) \leq \max_{x,y\in\Omega} \mathbb{P}_{x,y}(\tau_c > t), \tag{1.39}$$

where τ_c is defined as in the previous theorem.

Proof: By the definition of the stationary distribution π

$$\begin{aligned}
\|P^t(x,\cdot) - \pi\|_{TV} &= \max_{A\subset\Omega} |P^t(x,A) - \pi(A)| \\
&= \max_{A\subset\Omega} \Big| \sum_{y\in\Omega} \pi(y)(P^t(x,A) - P^t(y,A)) \Big| \\
&\leq \max_{A\subset\Omega} \sum_{y\in\Omega} \pi(y) |P^t(x,A) - P^t(y,A)| \\
&\leq \sum_{y\in\Omega} \pi(y) \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \\
&\leq \max_{y\in\Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}.
\end{aligned}$$

Taking also the maximum over all possible x's, we have finally

$$d(t) \leq \max_{x,y\in\Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}.$$

Putting together this and Theorem 1.15 gives the corollary. ■

We are now able to prove Proposition 1.9:

Proof (of Proposition 1.9): Let's consider two different starting states $x, y \in \Omega$ for the chain. By definition of $t_{mix}(c)$, we have both

$$\|P^T(x,\cdot) - \pi(\cdot)\|_{TV} \leq c \quad \text{and} \quad \|P^T(y,\cdot) - \pi(\cdot)\|_{TV} \leq c.$$

Hence, by triangular inequality, $\|P^T(x,\cdot) - P^T(y,\cdot)\|_{TV} \leq 2c$. As we have seen in Proposition 1.14 we can construct a particular coupling (X_T, Y_T) of the measures $P^T(x,\cdot)$ and $P^T(y,\cdot)$ such that $\mathbb{P}(X_T \neq Y_T) \leq 2c$. Now consider a new Markov chain on the same state space with transition matrix $Q := P^T$; in words this new process performs a step every T steps of the original chain P. The coupling we have described above guarantees that the probability that the two instances of the new chain starting from x and y have not coupled in one "Q-step" is at most $2c$. So, by induction, the probability that the two copies of the Q-chain have not coupled in k steps is less or equal to $(2c)^k$. By Corollary 1.16, Q is within variation distance ε from its stationary distribution after k steps if

$$(2c)^k \leq \varepsilon.$$

It follows that after $\lceil \log \varepsilon / \log (2c) \rceil$ steps the new chain is closer than ε to its stationary distribution. But Q and P have the same stationary distribution, and one Q-step is equivalent to T P-steps. Therefore,

$$t_{mix}(\varepsilon) \leq \left\lceil \frac{\log \varepsilon}{\log (2c)} \right\rceil T$$

for the original Markov chain. ■

Chapter 2

The Cut-off phenomenon

2.1 Cut-off

2.1.1 Main definition

Now that our knowledge of finite Markov chains is pretty good, we are ready to introduce the main subject of this work.

Take a sequence of Markov chains indexed by $n = 1, 2, ...$ with their state space $\Omega^{(n)}$; we can think that as n grows, the size of the Ω's grows. Suppose that each of these chains has its stationary distribution $\pi^{(n)}$ and its mixing time $t^{(n)}_{mix}(\varepsilon)$. We say that the sequence has a **cut-off** if, for any $0 < \varepsilon < \frac{1}{2}$,

$$\lim_{n\to\infty} \frac{t^{(n)}_{mix}(\varepsilon)}{t^{(n)}_{mix}(1-\varepsilon)} = 1. \tag{2.1}$$

In general, the time to reach a distance $(1-\varepsilon)$ from the stationary distribution is smaller than the time to reach a distance ε. We can write

$$t^{(n)}_{mix}(\varepsilon) - t^{(n)}_{mix}(1-\varepsilon) = \tau^{(n)} > 0.$$

What equation (2.1) is telling us is that, in the limit for $n \to \infty$, $\tau^{(n)}$ becomes negligible on a time-scale of $t^{(n)}_{mix}$. Since ε can be taken arbitrarily small, we are also saying that the time to pass from the maximum distance from π to an almost-0 distance is (relatively) very short.

Graphically, if we draw the function $d_n(t)$ (that is the $d(t)$ fot the chain n) and zoom out the picture on a time-scale of $t^{(n)}_{mix}$, we see that as n grows the function approaches a step-function with jump at $t^{(n)}_{mix}$ (see Figure 2.1.1).

Knowing the existence of the cut-off for some family of Markov chains can be very useful: if we want to approach the distribution $\pi^{(n)}$ (for example in a simulation), we know that we have at least to wait $t^{(n)}_{mix}$ steps but also that is quite useless to run the chain much longer. This nice property seems to appear in many 'natural' Markov chains; nevertheless, nowadays the family

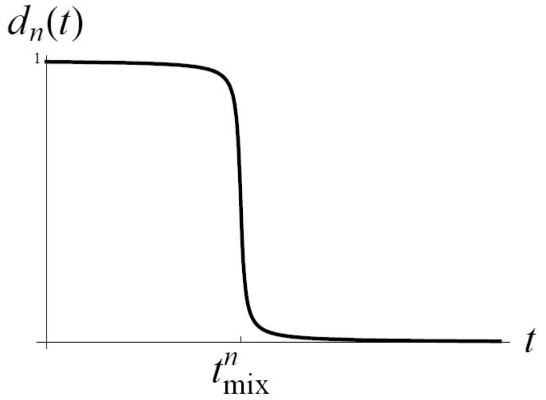

Figure 2.1: *For a sequence of chains with cut-off, the graph of $d_n(t)$ against t, zoomed on a time-scale of $t_{mix}^{(n)}$, approaches a step function as $n \to \infty$.*

of processes that are known to have a cut-off is pretty small, since verifying that a given sequence satisfies the definition can be quite hard. For this reason there are other weaker definitions for the cut-off, such as the pre-cutoff and the weak l^p-cut-off.

For example a sequence of Markov chains indexed by n on the state spaces Ω_n, with stationary distributions $\pi^{(n)}$ and mixing times $t_{mix}^{(n)}(\varepsilon)$, is said to show a **pre-cut-off** if

$$\sup_{0<\varepsilon<\frac{1}{2}} \limsup_{n\to\infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} < \infty. \tag{2.2}$$

In 2004 David Aldous showed a simple chain that has pre-cut-off but not cut-off.

2.1.2 An equivalent definition

Lemma 2.1. *A sequence of Markov chains exhibits a cut-off* $\Longleftrightarrow$

$$\Longleftrightarrow \; d_n(c\, t_{mix}^{(n)}) \xrightarrow{n\to\infty} \begin{cases} 1 & \text{if } c<1 \\ 0 & \text{if } c>1 \end{cases}$$

Proof: $\Leftarrow$) From the definition of limit, for any $\gamma > 0$ we can choose n big enough to make $d_n((1-\gamma)\, t_{mix}^{(n)})$ arbitrarily close to 1. It follows that for any $0 < \varepsilon < \frac{1}{2}$ we can choose n such that $d_n((1-\gamma)\, t_{mix}^{(n)}) > (1-\varepsilon)$. Hence, for such n,

$$t_{mix}^{(n)}(1-\varepsilon) \geq (1-\gamma) t_{mix}^{(n)}; \tag{2.3}$$

in the very same way we can choose n big enough such that

$$t_{mix}^{(n)}(\varepsilon) \leq (1+\gamma)t_{mix}^{(n)}. \tag{2.4}$$

Therefore exists n such that

$$1 \leq \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} \leq \frac{1+\gamma}{1-\gamma}, \tag{2.5}$$

and taking the limit for $\gamma \to 0$ (so that $n \to \infty$) we obtain the definition of cut-off.

$\Rightarrow$) From the definition of limit, fixing γ, there exists $\bar{n}$ such that $\forall n > \bar{n}$

$$\frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} \leq (1+\gamma) = c \tag{2.6}$$

so that

$$t_{mix}(\varepsilon) \leq c\, t_{mix}^{(n)}(1-\varepsilon) \leq c\, t_{mix}^{(n)}.$$

Therefore, from the definition of $d^{(n)}(\cdot)$ and taking the limit over n to cover all possible $0 < \varepsilon < \frac{1}{2}$,

$$\lim_{n\to\infty} d^{(n)}(c\, t_{mix}^{(n)}) \leq \varepsilon. \tag{2.7}$$

Letting $\varepsilon \to 0$ we have the case $c > 1$. The case $c < 1$ is completely analogous. ■

2.1.3 The cut-off window

We would like to be more precise to describe how long it takes for the chain to fall from a distance ~ 1 to a distance ~ 0. We say that a sequence of Markov chains has a cut-off with a **window** of size ω_n if $\omega_n = o(t_{mix}^{(n)})$ and

$$\lim_{\alpha\to\infty} \liminf_{n\to\infty} d_n(t_{mix}^{(n)} - \alpha\omega_n) = 1, \tag{2.8}$$

$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} d_n(t_{mix}^{(n)} + \alpha\omega_n) = 0. \tag{2.9}$$

Equivalently we can say that a sequence $\{\omega_n\}$ is a cut-off window for a family of chains $\left(X_t^{(n)}\right)$ if $\omega_n = o(t_{mix}^{(n)})$ and

$$t_{mix}^{(n)}(\varepsilon) - t_{mix}^{(n)}(1-\varepsilon) \leq c_\varepsilon \omega_n. \tag{2.10}$$

So with $\{\omega_n\}$ we know the size of the little interval around $t_{mix}^{(n)}$ in which the total variation distance of the chain from its stationary distribution collapses.

2.1.4 A necessary condition for the cut-off

Proposition 2.2. *For a sequence of irreducible, aperiodic and reversible Markov chains with spectral gaps* $\{\text{gap}^{(n)}\}$ *and mixing times* $\{t_{mix}^{(n)}\}$*, if*

$$\lim_{n\to\infty} t_{mix}^{(n)} \cdot \text{gap}^{(n)} < \infty,$$

then there is no pre-cut-off.

Proof: Recall the first bound of Theorem 1.13:

$$(t_{rel} - 1)\log\left(\frac{1}{2\varepsilon}\right) \leq t_{mix}(\varepsilon).$$

Dividing both sides by $t_{mix}^{(n)}$ we know from the hypothesis that there exist a constant $c > 0$ such that, $\forall n$ and $\forall \varepsilon \in (0, \frac{1}{2})$,

$$\frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} \geq \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}} \geq \frac{t_{rel}^{(n)} - 1}{t_{mix}^{(n)}} \log\left(\frac{1}{2\varepsilon}\right) \geq c\log\left(\frac{1}{2\varepsilon}\right).$$

Letting $\varepsilon \to 0$, the left hand side goes to ∞. ■

2.2 An example: the simple random walk on the hypercube

2.2.1 The model

Till the end of this chapter we will apply the theoretical definitions and results seen so far to a concrete simple example.

The n-dimensional **hypercube** is a graph whose vertices are the binary strings of length n taking values in $\{0, 1\}^n$. Obviously there are 2^n of such vertices. Any two points are connected by an edge if and only if they differ for just one coordinate (or bit). If we imagine the picture of this graph in the 3-dimensional space, we obtain a typical cube of size 1. Its generalisation in n dimensions is the hypercube.

Our random process is the **simple random walk** on this graph. Imagine a cat that lies on a vertex of the hypercube; every "second" the cat chooses at random, that is with equal probability, one of the n vertices connected by an edge to its current position and jumps immediately there. The process will be the random sequence of the sites visited by the cat.

It is of course a Markov chain: the cat jumps over the next site without taking into account its past moves, but only considering its present position. The transition matrix P of the process has $\frac{1}{n}$-entries in the places $P(u, v)$ where u and v are connected by an edge, and is 0 everywhere else. Besides, the chain is irreducible, since after at most n jumps the cat can arrive with positive probability everywhere.

As all the cats, also our random cat can be very lazy. In this case we define a **lazy** version of the process, in which the cat can decide (with probability $\frac{1}{2}$), to rest a bit and remain for one "second" on its position. In the transition matrix of the process $\frac{1}{2}$ will appear on the whole diagonal and all the $\frac{1}{n}$'s will be replaced by $\frac{1}{2n}$. The advantage in this case is that the process becomes also aperiodic, so that we can apply a lot of the theorems we have seen in the first chapter.

Note that, by the detailed balance equations (1.8), it is very simple to verify that the uniform measure is the stationary distribution for the chain.

Finally we can observe that this process has a particularity: the chain looks the same from any point in the state space Ω (the set of all vertices). That is, the possible kind of choices of the cat are always the same, doesn't matter on which vertex he is lying. Formally, for each couple $(x, y) \in \Omega \times \Omega$ there is a bijection $\phi = \phi_{(x,y)} : \Omega \to \Omega$ such that $\phi(x) = y$ and $P(z, w) = P(\phi(z), \phi(w))$, $\forall z, w \in \Omega$. Such kind of chains is called **transitive**. One obvious consequence of this property is that in order to calculate the total variation distance of the chain from its stationary distribution we can let the chain start from any point of Ω.

2.2.2 Mixing time

Let's start studying the mixing time of our lazy random walk on the hypercube. With a very simple and intuitive coupling we will be able to bound t_{mix} from above and this bound will be of the correct order in n up to constants.

This is the description of our coupling: take two copies of the chain starting from any two vertices of the hypercube. Choose at random one of the n coordinates and toss a fair coin: if it comes up heads, upload in *both* chains the chosen coordinate with a 0 bit, otherwise with a 1. This way, once the j-th coordinate has been chosen, the two chains will have the same bit in the j-th position forever.

Clearly, if all coordinate have been chosen at least once, the two chains proceed together. Then it is useful to define

$$\tau := \{\text{First time all the coordinates have been chosen at least once}\}.$$

Proposition 2.3. *$\forall c > 0$,*

$$\mathbb{P}\left(\tau > \lceil n \log n + cn \rceil\right) \le e^{-c}. \tag{2.11}$$

Proof: Define the events

$$A_i := \{\text{The } i\text{-th coordinate have not been chosen in the first } \lceil n \log n + cn \rceil \text{ steps}\}.$$

Then, $\forall i = 1, 2, ..., n$,

$$\mathbb{P}(A_i) = \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \leq e^{-\lceil \log n + c \rceil} \leq \frac{1}{n} e^{-c}.$$

Therefore

$$\mathbb{P}(\tau > \lceil n \log n + cn \rceil) = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) \leq e^{-c}.$$

■

Proposition 2.4. *For the simple lazy random walk on the n-dimensional hypercube*

$$t_{mix}(\varepsilon) \leq n \log n + n \log\left(\frac{1}{\varepsilon}\right). \tag{2.12}$$

Proof: From the coupling we have described above, from Corollary 1.16 and from equation (2.11) we obtain

$$d(n \log n + cn) \leq \mathbb{P}(\tau > \lceil n \log n + cn \rceil) \leq e^{-c},$$

and the conclusion follows immediately from the definition of mixing time.

■

The order $O(n \log n)$ we have found is actually the right order for the mixing time, but it can be improved by a constant factor of $\frac{1}{2}$ as it will be shown later in this chapter.

2.2.3 The Ehrenfest urn and the Hamming's weight

We want to introduce another process that at a first glance could appear totally unlinked from the original random walk. Imagine to have two urns, say Urn I and Urn II, in which are distributed n balls. At each step of the process we choose at random one of the n balls and transfer it from its current urn to the other. So, if we call Y_t the number of balls in Urn I at time t, the transition matrix of our chain is

$$P(j,k) = \begin{cases} \frac{n-j}{n} & \text{if } k = j+1 \\ \frac{j}{n} & \text{if } k = j-1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Y_t is a Markov chain on the state space $\Omega_E := \{0, 1, ..., n\}$ and, as one can see from the transition matrix, it has a drift towards the middle of this interval. Of course, if we leave a $\frac{1}{2}$ probability for the balls to stand still at each step, we obtain a lazy version of the chain. Thanks to the detailed balance equations it is very easy to verify that π_W, the stationary distribution for both these chains (lazy and not lazy), is a binomial distribution of parameters n and $\frac{1}{2}$, so that $\forall t$

$$\mathbb{E}_{\pi_W}[W_t] = \frac{n}{2} \tag{2.14}$$

and

$$Var_{\pi_W}(W_t) = \frac{n}{4}. \tag{2.15}$$

Which is the link between this chain, called the **Ehrenfest urn process**, and the random walk on the hypercube?

For a vertex of the hypercube, we can define its **Hamming's weight** as the number of 1's that appear in the string representation of such vertex:

$$W(x) = \sum_{j=1}^{n} x(j) \tag{2.16}$$

where $x(j) \in \{0, 1\}$. If X_t is the (not-lazy) random walk, then we can study its projection

$$W_t := W(X_t). \tag{2.17}$$

If $W_t = j$, then it increments by a unit amount if one of the $n - j$ 0-coordinates is flipped, while it decrements if one of the j 1-coordinates is chosen. It follows that W_t is again a Markov chain and that its transition probabilities are described as well by (2.13).

In this parallel the j-th coordinate of the random walk X_t can be thought as the j-th ball of the urns process: if the j-th bit is 1, then the ball is in Urn I, while if the bit is 0 the ball is in Urn II.

The following lemma provides another key connection between the two models. It will allow us to deal with the Ehrenfest urn instead of the more complicated random walk on the hypercube in order to demonstrate the upper bound for the cut-off.

Lemma 2.5. *Let X_t be the simple random walk on the hypercube and let $W_t := \sum_{j=1}^{n} X_t(j)$ be the Hamming's weight of the chain at each step (corresponding to an Ehrenfest urn chain). Then*

$$\|\mathbb{P}_{\bar{1}}(X_t \in \cdot) - \pi\|_{TV} = \|\mathbb{P}_n(W_t \in \cdot) - \pi_W\|_{TV}, \tag{2.18}$$

where $\mathbb{P}_{\bar{1}}$ says that we are starting the random walk by the vertex with all 1's coordinates and $\mathbb{P}_n$ says that we are starting the urns chain with all the balls in Urn I.

Proof: Define $\Omega_w := \{x \in \Omega : W(x) = w\}$. By symmetry, both $x \to \mathbb{P}_{\bar{1}}(X_t = x)$ and $x \to \pi(x)$ are constant functions over Ω_w. Then

$$\sum_{\substack{x\in\Omega:\\ W(x)=w}} |\mathbb{P}_{\bar{1}}(X_t = x) - \pi(x)| = \Big| \sum_{\substack{x\in\Omega:\\ W(x)=w}} \mathbb{P}_{\bar{1}}(X_t = x) - \pi(x) \Big|$$
$$= |\mathbb{P}_{\bar{1}}(W_t = x) - \pi_W(w)|,$$

where we used the fact that all the terms in the first summation are equal. Summing over all possible $w \in \{0, 1, ..., n\}$ and dividing by 2 we obtain equation (2.18). ■

2.2.4 Some other tools

In this section we will state some general results that will be useful for the proof of the cut-off for the random walk on the hypercube.

One way to produce a lower bound for the mixing time t_{mix} is to find a statistic (a real-valued function) f on Ω such that the distance between $f(X_t)$ and the distribution of f under the stationary distribution π can be bounded from below. In fact we have the following lemmas:

Lemma 2.6. *Let μ and ν be two probability measures on Ω and let $f : \Omega \to \Lambda$, with Λ a finite set. Setting, $\forall A \subset \Omega$,*

$$\mu f^{-1}(A) := \mu(\{x : f(x) \in A\}),$$

we have

$$\|\mu - \nu\|_{TV} \geq \|\mu f^{-1} - \nu f^{-1}\|_{TV}. \tag{2.19}$$

Proof:

$$\max_{B\subset\Lambda} |\mu f^{-1}(B) - \nu f^{-1}(B)| = \max_{B\subset\Lambda} |\mu(f^{-1}(B)) - \nu(f^{-1}(B))|$$
$$\leq \max_{A\subset\Omega} |\mu(A) - \nu(A)|.$$

■

Lemma 2.7. *Let $f : \Omega \to \Lambda$, with Λ a finite set, and let μ and ν two probability measures on Ω such that*

$$\mathbb{E}^{\nu}[f] - \mathbb{E}^{\mu}[f] \geq r\sigma_*,$$

for some $r > 0$, where

$$\mathbb{E}^{\mu}[f] = \sum_{x\in\Omega} \mu(x) f(x)$$

and

$$\sigma_* := \sqrt{\max\{Var_{\mu}(f), Var_{\nu}(f)\}}.$$

Then

$$\|\mu - \nu\|_{TV} \geq 1 - \frac{8}{r^2}. \tag{2.20}$$

Proof: Let

$$A := \left(\mathbb{E}^{\mu}[f] + \frac{r\sigma_*}{2}, \infty\right)$$

and apply Chebychev inequality:

$$\begin{aligned}
\mu f^{-1}(A) &= \mu(\{x : f(x) \geq \mathbb{E}^{\mu}[f] + \frac{r\sigma_*}{2}\}) \\
&\leq \mu(\{x : |f(x) - \mathbb{E}^{\mu}[f]| \geq \frac{r\sigma_*}{2}\}) \\
&\leq \frac{4}{r^2}
\end{aligned}$$

and

$$\begin{aligned}
\nu f^{-1}(A) &= \nu(\{x : f(x) \geq \mathbb{E}^{\mu}[f] + \frac{r\sigma_*}{2}\}) \\
&\geq \nu(\{x : f(x) \geq \mathbb{E}^{\nu}[f] - r\sigma_* + \frac{r\sigma_*}{2}\}) \\
&\geq \nu(\{x : |f(x) - \mathbb{E}^{\nu}[f]| \leq \frac{r\sigma_*}{2}\}) \\
&\geq 1 - \frac{4}{r^2}.
\end{aligned}$$

Therefore, thanks to Lemma 2.6, we have

$$\begin{aligned}
\|\mu - \nu\|_{TV} &\geq \|\mu f^{-1} - \nu f^{-1}\|_{TV} = \sup_{B \subset \Lambda} |\mu f^{-1}(B) - \nu f^{-1}(B)| \\
&\geq |\mu f^{-1}(A) - \nu f^{-1}(A)| \geq 1 - \frac{4}{r^2} - \frac{4}{r^2} \\
&= 1 - \frac{8}{r^2}.
\end{aligned}$$

■

Going back to the problem of knowing how long we have to wait to refresh the coordinates in the random walk on the hypercube, the last lemma of this section gives a good amount of informations.

Lemma 2.8. *Consider the simple random walk on the n-dimensional hypercube. Let*

$$I_j(t) := \chi_{\{j\text{-th coordinate has not been chosen up to time } t\}}$$

and

$$R_t := \sum_{j=1}^{n} I_j(t) = \#\{\text{coordinates not refreshed at time } t\}.$$

Therefore the $I_j(t)$'s are negatively correlated and if

$$p := \left(1 - \frac{1}{n}\right)^t,$$

then, $\forall t > 0$,

$$\mathbb{E}[R_t] = np \tag{2.21}$$

and

$$Var(R_t) \leq np(1-p) \leq \frac{n}{4}. \tag{2.22}$$

Proof: Since

$$\mathbb{P}(\{j\text{-th coordinate has not been chosen up to time } t\}) = p,$$

we have that the $I_j(t)$'s are Bernoulli random variables of parameter p; then

$$\mathbb{E}[I_j(t)] = p,$$
$$Var(I_j(t)) = p - p^2 = p(1-p)$$

and (2.21) follows immediately.

For $k \neq j$ we have

$$\mathbb{E}[I_j(t) I_k(t)] = \left(1 - \frac{2}{n}\right)^t;$$

therefore

$$Cov(I_j(t), I_k(t)) = \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{1}{n}\right)^{2t} \leq 0, \tag{2.23}$$

and (2.22) is a direct consequence, keeping in mind that $\forall q \in [0,1]$

$$q(1-q) \leq \frac{1}{2}.$$

■

2.2.5 Lower bound for the cut-off

Proposition 2.9. *For the lazy simple random walk on the n-dimensional hypercube*

$$d\left(\frac{1}{2} n \log n - \alpha n\right) \geq 1 - 8e^{-2\alpha+1}. \tag{2.24}$$

Proof: We would like to apply Lemma 2.7 with $f(\cdot) = W(\cdot)$, $\mu = \pi$ and $\nu = \delta_{\bar{1}}$ (that is the random walk starting by the vertex with all the coordinates equal to 1: $X_0 = (1, 1, ..., 1)$). Let's evaluate σ_*.

First of all note that by (2.15) we have

$$Var_\pi(W(X_t)) = Var_{\pi_W}(W_t) = \frac{n}{4}. \tag{2.25}$$

Then, using the notation of Lemma 2.8, call R_t the number of coordinates that have not been selected up to time t. Starting by the $\bar{1}$-configuration, the distribution of $W(X_t)$ given $R_t = r$ is the distribution of the random variable $B + r$, where B is a binomial of parameters $(n-r)$ and $\frac{1}{2}$. Then

$$\mathbb{E}_{\bar{1}}\left[W(X_t)|R_t\right] = R_t + \frac{n-R_t}{2} = \frac{1}{2}(R_t + n); \tag{2.26}$$

taking another expectation over the possible values of R_t and using equation (2.21) we obtain

$$\mathbb{E}_{\bar{1}}\left[W(X_t)\right] = \frac{n}{2}\left[1 + \left(1 - \frac{1}{n}\right)^t\right]. \tag{2.27}$$

Plugging in equation (2.26) in the identity given by the well known 'total variation formula'

$$Var_{\bar{1}}(W(X_t)) = Var(\mathbb{E}_{\bar{1}}\left[W(X_t)|R_t\right]) + \mathbb{E}\left[Var_{\bar{1}}(W(X_t)|R_t)\right] \tag{2.28}$$

and remembering the nature of $W(X_t)$ given R_t, we have

$$Var_{\bar{1}}(W(X_t)) = \frac{1}{4}Var(R_t) + \frac{1}{4}\left[n - \mathbb{E}\left[R_t\right]\right]. \tag{2.29}$$

Since R_t is the summation of indicator functions negatively correlated

$$\begin{aligned} Var(R_t) &= \mathbb{E}\Big[\Big(\sum_j I_j\Big)^2\Big] - \mathbb{E}^2\Big[\sum_j I_j\Big] \\ &= \sum_{j\neq k}\mathbb{E}\left[I_j I_k\right] + \sum_j \mathbb{E}\left[I_j{}^2\right] - \sum_{j\neq k}\mathbb{E}\left[I_j\right]\mathbb{E}\left[I_k\right] - \sum_j \mathbb{E}^2\left[I_j\right] \\ &\leq \sum_j \mathbb{E}\left[I_j\right] - \sum_j \mathbb{E}^2\left[I_j\right] \\ &\leq \mathbb{E}\left[R_t\right], \end{aligned}$$

so

$$Var_{\bar{1}}(W(X_t)) \leq \frac{1}{4}np + \frac{1}{4}[n - np] = \frac{n}{4}.$$

This means that

$$\sigma_* = \frac{\sqrt{n}}{2}. \tag{2.30}$$

Therefore

$$\begin{aligned} \left|\mathbb{E}_{\bar{1}}\left[W(X_t)\right] - \mathbb{E}_{\pi}\left[W(X_t)\right]\right| &= \frac{n}{2}\left(1 - \frac{1}{n}\right)^t \\ &= \sigma_*\sqrt{n}\left(1 - \frac{1}{n}\right)^t \\ &= \sigma_* e^{-t\left(-\log\left(1-\frac{1}{n}\right)\right)} e^{\frac{\log n}{2}} \\ &\geq \sigma_* e^{-\frac{t}{n}\left(1+\frac{1}{n}\right)+\frac{\log n}{2}} \end{aligned}$$

where in the last inequality we've used the fact that $\log(1-x) \geq -x - x^2$, $\forall 0 \leq x \leq 1$.

From Lemma 2.7 we have

$$\|P^t(\bar{1}, \cdot) - \pi\|_{TV} \geq 1 - 8e^{\frac{2t}{n}\left(1+\frac{1}{n}\right) - \log n}$$

and since

$$t_n := \left[1 - \frac{1}{n+1}\right]\left[\frac{1}{2} n \log n - \left(\alpha - \frac{1}{2}\right) n\right] \geq \frac{1}{2} n \log n - \alpha n$$

we finally obtain

$$\begin{aligned} d\left(\frac{1}{2} n \log n - \alpha n\right) &\geq d(t_n) \\ &\geq 1 - 8e^{\frac{2t_n}{n}\left(1+\frac{1}{n}\right) - \log n} \\ &= 1 - 8e^{-2\alpha+1}. \end{aligned} \tag{2.31}$$

■

2.2.6 Upper bound for the cut-off

As we said before, the result of Section 2.2.2 can be improved by a constant factor, completing the proof of the existence of the cut off for our Markov chain.

Proposition 2.10. *For the lazy simple random walk on the n-dimensional hypercube there exists a constant $c > 0$ such that*

$$d\left(\frac{1}{2} n \log n + \alpha n\right) \leq \frac{c}{\sqrt{\alpha}}. \tag{2.32}$$

Proof: Since the chain is transitive, we have, recalling Lemma 2.5,

$$d(t) = \|\mathbb{P}_{\bar{1}}(X_t \in \cdot) - \pi\|_{TV} = \|\mathbb{P}_n(W_t \in \cdot) - \pi_W\|_{TV}, \tag{2.33}$$

so that it will be sufficient to bound the right hand side of (2.33). We will use again the powerful coupling method to do that.

Let's build two copies of the Ehrenfest urn model, W_t and Z_t, starting from different points of Ω_W, $W_0 = w$ and $Z_t = z$ with $z \geq w$ without loss of generality. At each step we throw a fair coin in order to decide which of the two copies will perform a move according to the probabilities described in table (2.13). This way, looking separately at W_t and Z_t, we will see two copies of the lazy Ehrenfest urn process. Besides, once the two chains meet, we can force them to stay together forever.

Define

$$D_t := Z_t - W_t, \tag{2.34}$$

the difference of the number of balls in Urn I in the two copies; by construction $D_t \geq 0$, $\forall t \geq 0$. Let also

$$\tau := \min\{t \geq 0 : Z_t = W_t\} \tag{2.35}$$

be the first time the two copies meet.

As long as $\tau > t$,

$$D_{t+1} - D_t = \begin{cases} 1 & \text{with prob. } \frac{1}{2}\left(1 - \frac{Z_t}{n}\right) + \frac{1}{2}\frac{W_t}{n} \\ -1 & \text{with prob. } \frac{1}{2}\frac{Z_t}{n} + \frac{1}{2}\left(1 - \frac{W_t}{n}\right). \end{cases} \tag{2.36}$$

Therefore, if $\tau > t$,

$$\mathbb{E}_{z,w}\left[D_{t+1} - D_t | Z_t = z_t, W_t = w_t\right] = -\frac{z_t - w_t}{n} = -\frac{d_t}{n}. \tag{2.37}$$

From this, the fact that $\chi_{\{\tau>t\}}$ depends only on the history of the chain up to time t and from the Markov property (1.1), we have

$$\mathbb{E}_{z,w}\left[\chi_{\{\tau>t\}} D_{t+1} | Z_0, ..., Z_t, W_0, ..., W_t\right] = \left(1 - \frac{1}{n}\right) \chi_{\{\tau>t\}} D_t. \tag{2.38}$$

Taking the expectation over all possible paths of Z and W

$$\mathbb{E}_{z,w}\left[\chi_{\{\tau>t\}} D_{t+1}\right] = \left(1 - \frac{1}{n}\right) \mathbb{E}_{z,w}\left[\chi_{\{\tau>t\}} D_t\right] \tag{2.39}$$

and since, $\forall t$,

$$\chi_{\{\tau>t+1\}} \leq \chi_{\{\tau>t\}}$$

we have

$$\mathbb{E}_{z,w}\left[\chi_{\{\tau>t+1\}} D_{t+1}\right] \leq \left(1 - \frac{1}{n}\right) \mathbb{E}_{z,w}\left[\chi_{\{\tau>t\}} D_t\right]. \tag{2.40}$$

Iterating (2.40)

$$\begin{aligned} \mathbb{E}_{z,w}\left[\chi_{\{\tau>t\}} D_t\right] &\leq \left(1 - \frac{1}{n}\right)^t (z - w) \\ &\leq n e^{-\frac{t}{n}}. \end{aligned} \tag{2.41}$$

Now note that if $\tau > t$ the increments $D_{t+1} - D_t$ tend to be negative; in fact, the probabilities in (2.36) and equation (2.37) say that at each step D_t increases of one unit with probability not bigger than $\frac{1}{2} - \frac{1}{2n}$ and decreases of one unit with probability at least $\frac{1}{2} + \frac{1}{2n}$. Therefore it is possible to couple D_t with a symmetric random walk S_t on the state space $\Omega_S := \mathbb{N}_0$ that has probability $\frac{1}{2} - \frac{1}{2n}$ to go either to the left or to the right and that is slightly lazy (it stands still with probability $\frac{1}{n}$). We can suppose $S_0 = z - w$ and force S_t to dominate D_t in the sense that whenever S_t performs a step to the left, we oblige D_t to do the same; this way $\forall t \leq \tau$, $S_t \geq D_t$.

Call $\tilde{\tau} := \min\{t \geq 0 : S_t = 0\}$. Then, for what we said, $\tau \leq \tilde{\tau}$ and by a general result for the simple random walks that will be proved in Corollary 2.12 at the end of this section, $\exists c > 0$ such that for $k \geq 0$

$$\mathbb{P}_{z,w}\left(\tau > u\right) \leq \mathbb{P}_{z-w}\left(\tilde{\tau} > u\right) \leq \frac{c(z-w)}{\sqrt{u}}. \tag{2.42}$$

Therefore

$$\mathbb{P}_{z,w}\left(\tau > s+u | D_0, D_1, ..., D_s\right) = \chi_{\{\tau > s\}} \mathbb{P}_{D_s}\left(\tau > u\right) \leq \chi_{\{\tau > s\}} \frac{cD_s}{\sqrt{u}}; \tag{2.43}$$

taking the expectation and applying (2.41)

$$\mathbb{P}_{z,w}\left(\tau > s+u\right) \leq \frac{cne^{-\frac{s}{n}}}{\sqrt{u}}. \tag{2.44}$$

Choosing $s = \frac{1}{2} n \log n$ and $u = \alpha n$ we finally obtain

$$\mathbb{P}_{z,w}\left(\tau > \frac{1}{2} n \log n + \alpha n\right) \leq \frac{c}{\sqrt{\alpha}}. \tag{2.45}$$

By Corollary 1.16 we have the thesis. ∎

There is only left to prove the general result on the random walks we used in (2.42).

Theorem 2.11. *Let $\{Z_i\}$ be i.i.d. integer-valued random variables with $\mathbb{E}[Z_i] = 0$ and $Var(Z_i) = \sigma^2$, $\forall i$. If we define $X_t := \sum_{i=1}^{t} Z_i$, then*

$$\mathbb{P}\left(X_t \neq 0 \textit{ for } 1 \leq t \leq r\right) \leq \frac{4\sigma}{\sqrt{r}}. \tag{2.46}$$

Proof: For $I \subset \mathbb{Z}$ let

$$L_r(I) := \{t \in \{0, 1, ..., r\} : X_t \in I\}$$

be the set of times up to r in which $X_t \in I$. Then let

$$A_r := \{t \in L_r(0) : X_{t+u} \neq 0 \text{ for } 1 \leq u \leq r\}$$

be the set of times t in $L_r(0) = L_r(\{0\})$ after which the walk doesn't touch 0 for other r steps (clearly $|A_r| \leq 1$).

Since the future of the walk after visiting 0 doesn't depend on what happened before,

$$\mathbb{P}\left(t \in A_r\right) = \mathbb{P}\left(t \in L_r(0)\right) \alpha_r$$

where

$$\alpha_r := \mathbb{P}_0\left(X_t \neq 0, t = 1, 2, ..., r\right).$$

Summing over t gives

$$\begin{aligned}
1 \geq \mathbb{E}\left[|A_r|\right] &= \mathbb{E}\left[\sum_{t=0}^{r} \chi_{\{t \in A_r\}}\right] = \sum_{t=0}^{r} \mathbb{P}\left(t \in A_r\right) \\
&= \sum_{t=0}^{r} \mathbb{P}\left(t \in L_r(0)\right) \alpha_r = \mathbb{E}\left[\sum_{t=0}^{r} \chi_{\{t \in L_r(0)\}}\right] \alpha_r \\
&= \mathbb{E}\left[|L_r(0)|\right] \alpha_r. \qquad (2.47)
\end{aligned}$$

It remains only to estimate $\mathbb{E}\left[|L_r(0)|\right]$ from below. By Chebychev inequality (keeping in mind that $Var(X_t) = t\sigma^2$)

$$\mathbb{P}\left(|X_t| \geq \sigma\sqrt{r}\right) = \mathbb{P}\left(|X_t - 0| \geq \sigma\sqrt{t}\frac{\sqrt{r}}{\sqrt{t}}\right) \leq \frac{t}{r}.$$

Taking $I := (-\sigma\sqrt{r}, \sigma\sqrt{r})$,

$$\mathbb{E}\left[|L_r(I^c)|\right] = \mathbb{E}\left[\sum_{t=0}^{r} \chi_{\{t \in L_r(I^c)\}}\right] \leq \sum_{t=0}^{r} \frac{t}{r} = \frac{r+1}{2},$$

whence

$$\mathbb{E}\left[|L_r(I)|\right] = r + 1 - \mathbb{E}\left[|L_r(I^c)|\right] > \frac{r}{2}.$$

Furthermore, for any $v \neq 0$,

$$\begin{aligned}
\mathbb{E}\left[|L_r(v)|\right] &= \mathbb{E}\left[\sum_{t=0}^{r} \chi_{\{X_t = v\}}\right] = \mathbb{E}\left[\sum_{t=\tau_v}^{r} \chi_{\{X_t = v\}}\right] \leq \mathbb{E}_v\left[\sum_{t=0}^{r} \chi_{\{X_t = v\}}\right] \\
&= \mathbb{E}_0\left[\sum_{t=0}^{r} \chi_{\{X_t = 0\}}\right],
\end{aligned}$$

where for the inequality we have used the Markov property, which says that the chain after τ_v has the same distribution of the chain started from v.

Thus

$$\frac{r}{2} \leq \mathbb{E}\left[|L_r(I)|\right] \leq 2\sigma\sqrt{r}\mathbb{E}\left[|L_r(0)|\right],$$

that in conjunction with (2.47) proves the theorem. ■

Corollary 2.12. *Let X_t ba a simple random walk on $\mathbb{Z}$ with probability $0 \leq p \leq \frac{1}{2}$ of remaining in its current position at each step. Then*

$$\mathbb{P}_k\left(\tau_0 > r\right) \leq \frac{8|k|}{\sqrt{r}}. \qquad (2.48)$$

Proof: By conditioning on the first move of the walk and using the fact that the distribution of the walk is symmetric about 0, for $r \geq 1$

$$\begin{aligned}\mathbb{P}_0\left(\tau_0^+ > r\right) &= \frac{1-p}{2}\mathbb{P}_1\left(\tau_0 > r-1\right) + \frac{1-p}{2}\mathbb{P}_{-1}\left(\tau_0 > r-1\right) \\ &\geq \frac{1}{2}\mathbb{P}_1\left(\tau_0 > r-1\right) \qquad (2.49)\end{aligned}$$

where τ_0^+ indicates the first time the walk hits 0 *after* time $t = 0$. Note that the event {the walk hits k before 0 and then for r steps doesn't touch 0} is contained in the event {the walk doesn't touch 0 for $r-1$ steps}. Then, using both (2.46) and (2.49), and reminding that for this kind of random walks $\sigma^2 \leq 1$, we have

$$\mathbb{P}_1\left(\tau_k < \tau_0\right)\mathbb{P}_k\left(\tau_0 > r\right) \leq \mathbb{P}_1\left(\tau_0 > r-1\right) \leq 2\mathbb{P}_0\left(\tau_0^+ > r\right) \leq \frac{8}{\sqrt{r}}. \qquad (2.50)$$

It is well known (see the *gambler's ruin problem*, e.g. Proposition 2.1 in [17]) that

$$\mathbb{P}_1\left(\tau_k < \tau_0\right) = \frac{1}{k},$$

and the thesis follows immediately. ■

2.2.7 Conclusion

Theorem 2.13. *The lazy random walk on the n-dimensional hypercube has a cut-off at $\frac{1}{2}n\log n$ with a window of size n.*

Proof: Propositions 2.9 and 2.10. ■

Finally, using a lot of interesting techniques, we managed to prove the existence of the *cut-off* for our random walk. We have seen that the stationary distribution for the chain is the uniform measure over all the vertices of the hypercube and that to be sufficiently close to this measure we have to run the chain for about $\frac{1}{2}n\log n$ steps, that is the correct order for t_{mix}. If we wait less than this time, we are still able some way to recognize from which position we started and the uniform distribution is not well approximated, while waiting more is quite useless. In particular, these 'less' and 'more' are quantifiable: they are exactly described by the $O(n)$ of the window size.

Chapter 3

Birth and death processes

3.1 Birth and death chains

3.1.1 The models

A **Birth and death chain** on $\Omega_n := \{0, 1, ..., n\}$ is a Markov chain P such that $P(x, y) = 0$ unless $|x - y| \le 1$. Write

$$b(x) := P(x, x+1) \qquad \text{for } x = 0, 1, ..., n-1, \tag{3.1}$$

$$r(x) := P(x, x) \qquad \forall x \in \Omega, \tag{3.2}$$

$$d(x) := P(x, x-1) \qquad \text{for } x = 1, 2, ..., n \tag{3.3}$$

and set for convention $d(0) = 0$ and $b(n) = 0$.

Two important classes of Birth and death chains are those with absorbing walls, that is $b(0) = 0$ and $d(n) = 0$ (once we arrive at the extreme points we stay there forever), and with repulsive, or partially repulsive, walls, that is with $b(0) > 0$ and $d(n) > 0$.

Here we are interested in irreducible chains, so that we will assume that $b(x) > 0$ for $x \in [0, n-1]$ and $d(x) > 0$ for $x \in [1, n]$. The stationary distribution for this kind of chains is

$$\pi(x) := \frac{1}{Z} \prod_{y=1}^{x} \frac{b(y-1)}{d(y)}, \tag{3.4}$$

where Z is a normalizing constant. In fact with this measure the detailed balance holds:

$$\begin{aligned} \pi(x)P(x, x+1) &= \frac{1}{Z} \prod_{y=1}^{x} \frac{b(y-1)}{d(y)} b(x) \\ &= \frac{1}{Z} \prod_{y=1}^{x+1} \frac{b(y-1)}{d(y)} d(x+1) = \pi(x+1)P(x+1, x). \end{aligned}$$

Irreducibility guarantees also that λ_2, the second greatest eigenvalue of the transition matrix, is strictly smaller than 1. Besides, allowing at least one state $x \in \Omega_n$ to have $r(x) > 0$, we avoid the problem of periodicity. In this case we know that -1 is not an eigenvalue of the transition matrix.

Given $0 < \varepsilon < 1$ we define the **quantile states** of the chain as

$$Q(\varepsilon) := \min \left\{ k \in \Omega_n : \sum_{j=0}^{k} \pi(j) \geq \varepsilon \right\} \tag{3.5}$$

and symmetrically

$$\tilde{Q}(\varepsilon) := \max \left\{ k \in \Omega_n : \sum_{j=k}^{n} \pi(j) \geq \varepsilon \right\}. \tag{3.6}$$

For the sake of simplicity we will assume that our chain verify $\forall 0 < \varepsilon < 1$

$$Q(\varepsilon) = \tilde{Q}(1 - \varepsilon).$$

Even if it is not generally true for all $\varepsilon \in (0, 1)$, we note that only at most n values of the parameter can break this rule for a chain with n states. Therefore for any countable family of chains we can remove a countable set of such critical values of ε saving the above equality.

3.1.2 Bound of the mixing time

In this section we collect a series of general results for Birth and death chains that will be useful in the next sessions.

Lemma 3.1. *For any lazy irreducible Birth and death chain, for any* $0 < \varepsilon < 1$ *and* $t \geq 0$

$$\|P^t(0, \cdot) - \pi\|_{TV} \leq \mathbb{P}_0 \left(\tau_{Q(1-\varepsilon)} > t \right) + \varepsilon. \tag{3.7}$$

Furthermore, for all $k \in \Omega_n$,

$$\|P^t(k, \cdot) - \pi\|_{TV} \leq \mathbb{P}_k \left(\max\{\tau_{Q(\varepsilon)}, \tau_{Q(1-\varepsilon)}\} > t \right) + 2\varepsilon. \tag{3.8}$$

Proof: We start a coupling $(X_t, \tilde{X}_t)$ of the chain such that $X_0 = 0$ and $\tilde{X}_0 \sim \pi$. The two copies evolve according to the following rule: at each step a fair coin is tossed; if it comes up heads, we perform a not-lazy move of X_t, otherwise we perform a not-lazy move of $\tilde{X}_t$. Once they meet, they proceed together. This way both chains individually act as the original lazy chain. Observe that the two chains never cross each other, since they have to meet before doing that.

Call $\tau_{Q(1-\varepsilon)}$ the first time the chain X_t hits $Q(1-\varepsilon)$ and notice that $\tilde{X}_{\tau_{Q(1-\varepsilon)}}$ is distributed according to π (since this is true at any time). Therefore, by the definition of $Q(1-\varepsilon)$,

$$\mathbb{P}_{0,\mu}\left(X_{\tau_{Q(1-\varepsilon)}} \geq \tilde{X}_{\tau_{Q(1-\varepsilon)}}\right) \geq 1-\epsilon.$$

This means that with probability at least $1-\varepsilon$ the two chains have coalesced before $\tau_{Q(1-\varepsilon)}$. By Corollary 1.16 we have

$$\begin{aligned}
\|P^t(0,\cdot)-\pi\|_{TV} &\leq \mathbb{P}_{0,\pi}\left(X_t \neq \tilde{X}_t\right) \\
&= 1-\mathbb{P}_{0,\pi}\left(X_t = \tilde{X}_t | t > \tau_{Q(1-\varepsilon)}\right)\mathbb{P}_0\left(t > \tau_{Q(1-\varepsilon)}\right) \\
&\leq 1-(1-\varepsilon)(1-\mathbb{P}_0\left(t < \tau_{Q(1-\varepsilon)}\right)) \\
&\leq \mathbb{P}_0\left(\tau_{Q(1-\varepsilon)} > t\right)+\varepsilon. \qquad (3.9)
\end{aligned}$$

The same argument can be used for X_t starting in k. In fact, the same coupling gives

$$\mathbb{P}_{k,\mu}\left(X_{\tau_{Q(\varepsilon)}} \leq \tilde{X}_{\tau_{Q(\varepsilon)}}\right) \geq 1-\epsilon$$

and

$$\mathbb{P}_{k,\mu}\left(X_{\tau_{Q(1-\varepsilon)}} \geq \tilde{X}_{\tau_{Q(1-\varepsilon)}}\right) \geq 1-\epsilon.$$

Therefore the probability that the two copies meet between $\tau_{Q(\varepsilon)}$ and $\tau_{Q(1-\varepsilon)}$ is at least $1-2\varepsilon$, giving the thesis as before. ■

Corollary 3.2. *If X_t is an irreducible lazy Birth and death chain on Ω_n, then, for any $0<\varepsilon<\frac{1}{16}$,*

$$t_{mix} \leq 16\max\left\{\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right], \mathbb{E}_n\left[\tau_{Q(\varepsilon)}\right]\right\}. \qquad (3.10)$$

Proof: Take two points $x, y \in \Omega_n$. Clearly at least one of the endpoints $s \in \{0,1\}$ verifies

$$\mathbb{E}_s\left[\tau_y\right] \geq \mathbb{E}_x\left[\tau_y\right] \qquad (3.11)$$

(that is, to go from one endpoint to y we have to cross x). Denoting with R the right hand side of (3.10) we have

$$\begin{aligned}
\mathbb{P}_x\left(\max\{\tau_{Q(\varepsilon)}, \tau_{Q(1-\varepsilon)}\} \geq R\right) &\leq \mathbb{P}_x\left(\tau_{Q(\varepsilon)} \geq R\right)+\mathbb{P}_x\left(\tau_{Q(1-\varepsilon)} \geq R\right) \\
&\leq \frac{\mathbb{E}_s\left[\tau_{Q(\varepsilon)}\right]}{R}+\frac{\mathbb{E}_{s'}\left[\tau_{Q(1-\varepsilon)}\right]}{R} \\
&\leq \frac{1}{16}+\frac{1}{16}=\frac{1}{8}
\end{aligned}$$

where we have used together Markov's inequality and (3.11). The proof now follows directly from (3.8). ■

3.1.3 Bound for the variance of the $Q(1-\varepsilon)$-hitting time

We have established that the order of the mixing time is at most the bigger expected value of the hitting times of $Q(1-\varepsilon)$ starting from 0 and of $Q(\varepsilon)$ starting from n. Without loss of generality we will assume from now that the first one is bigger than the second, so that we can simplify (3.10) as

$$t_{mix} \leq 16\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right], \text{ for } 0 < \varepsilon < \frac{1}{16}. \tag{3.12}$$

A key element for our work is a result of Karlin and McGregor [14], proved in its discrete version by Fill [11], which represents hitting times for Birth and death chains as a sum of independent exponential variables.

Theorem 3.3. *Consider a Birth and death chain on $\Omega_m := \{0, 1, ..., m\}$ and suppose that m is an absorbing state ($r(m) = 1$), while for all other states $x \in \Omega_m$, $b(x) > 0$ and $d(x) > 0$ (except for $d(0) = 0$). Then the probability generating function for the absorption time in m is*

$$f(u) := \prod_{j=0}^{m-1}\left[\frac{(1-\lambda_j)u}{1-\lambda_j u}\right] \tag{3.13}$$

where the λ_j's are the m non-unit eigenvalues of the transition matrix P. Furthermore, if P has non-negative eigenvalues, then the absorption time in m is distributed as the sum of m independent geometric random variables whose failure probabilities are the non-unit eigenvalues of P.

Since we are interested in the hitting time of certain states (namely $Q(1-\varepsilon)$) starting from 0, it is clearly equivalent to consider chains with the target state as an absorbing end point and this is just what the above theorem deals with.

Lemma 3.4. *Let (X'_t) be a lazy Birth and death chain on $\Omega_m := \{0, 1, ..., m\}$, with m absorbing state and $b(x), d(x) > 0$ for all the other $x \in \Omega_m$ (except for $d(0) = 0$). Let gap_m denote its spectral gap. Then*

$$Var_0(\tau_m) \leq \frac{\mathbb{E}_0\left[\tau_m\right]}{\text{gap}_m}. \tag{3.14}$$

Proof: Call $\lambda_0 \geq \lambda_1 \geq ... \geq \lambda_{m-1}$ the m non-unit eigenvalues of the transition matrix of (X'_t). By Proposition 1.11 we know that all the eigenvalues are positive, so that we can use the second part of Theorem 3.3 and consider τ_m as the sum of m independent geometric random variables of parameters $(1-\lambda_j)$, $j = 0, ..., m-1$. Therefore

$$\mathbb{E}_0\left[\tau_m\right] = \sum_{j=0}^{m-1}\frac{1}{1-\lambda_j}, \quad Var_0(\tau_m) = \sum_{i=0}^{m-1}\frac{\lambda_j}{(1-\lambda_j)^2}. \tag{3.15}$$

Since $\lambda_0 \geq \lambda_i$, $\forall i = 1, ..., m-1$, we have

$$Var_0(\tau_m) \leq \frac{1}{1-\lambda_0} \sum_{j=0}^{m-1} \frac{1}{\lambda_j} = \frac{\mathbb{E}_0\left[\tau_m\right]}{\text{gap}_m}$$

as required. ■

As we said before, the hitting time of a state in our original chain is distributed as the hitting time of that state in a chain where it is an absorbing end-point. Anyway we would like to bound the variance of $\tau_{Q(1-\varepsilon)}$ with objects of the only original chain.

Proposition 3.5. *Let (X_t) be a lazy irreducible Birth and death chain on Ω_n and call* gap *its spectral gap. For $0 < \varepsilon < 1$*

$$Var_0(\tau_{Q(1-\varepsilon)}) \leq \frac{\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]}{\varepsilon \cdot \text{gap}}. \tag{3.16}$$

Proof: The proof of the proposition is straight obtained by plugging the result of next lemma into (3.14). ■

Lemma 3.6. *Let X(t) be a lazy irreducible Birth and death chain on $\Omega_n := \{0, 1, ..., n\}$ and denote with* gap *its spectral gap. Fix $\varepsilon \in (0,1)$ and let $m := Q(1-\varepsilon)$. Consider now the lazy Birth and death chain on Ω_m with absorbing state in m and call* gap_m *its spectral gap. Then*

$$\text{gap}_m \geq \varepsilon \cdot \text{gap}. \tag{3.17}$$

Proof: By the representation of the spectral gap given in Lemma 1.12 we know that

$$\text{gap} = \min_{\substack{\tilde{f} \not\equiv 0 \\ \mathbb{E}_\pi[\tilde{f}]=0}} \frac{\langle (Id-P)\tilde{f}, \tilde{f}\rangle_\pi}{\langle \tilde{f}, \tilde{f}\rangle_\pi} = \min_{\substack{\tilde{f} \not\equiv 0 \\ \mathbb{E}_\pi[\tilde{f}]=0}} \frac{1}{2} \frac{\sum_{i,j} (\tilde{f}(i) - \tilde{f}(j))^2 P(i,j)\pi(i)}{\sum_i \tilde{f}(i)^2 \pi(i)}. \tag{3.18}$$

Note that gap_m is $(1-\theta)$, where θ is the largest eigenvalue of P_m, the principal sub-matrix on the first m rows and columns, indexed by $\{0, 1, ..., m-1\}$. By irreducibility of (X_t) it follows that P_m is strictly sub-stochastic. Besides, the reversibility of (X_t) implies that P_m is a symmetric operator on $\mathbb{R}^m$ with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$, that is $\langle P_m x, y\rangle_\pi = \langle x, P_m y\rangle_\pi$ for every $x, y \in \mathbb{R}^m$. The Rayleigh-Ritz formula (see e.g. Theorem XIII.1 in [20]) gives

$$\theta = \max_{\substack{x \in \mathbb{R}^m \\ x \neq 0}} \frac{\langle P_m x, x\rangle_\pi}{\langle x, x\rangle_\pi}.$$

Therefore

$$\text{gap}_m = 1-\theta = \min_{\substack{f\neq 0\\ f(k)=0\forall k\geq m}} \frac{\sum_{i=0}^n (f(i) - \sum_{j=0}^n P(i,j)f(j))f(i)\pi(i)}{\sum_{i=0}^n f(i)^2\pi(i)}$$
$$= \min_{\substack{f\neq 0\\ f(k)=0\forall k\geq m}} \frac{1}{2}\frac{\sum_{i,j=0}^n (f(i)-f(j))^2 P(i,j)\pi(i)}{\sum_{i=0}^n f(i)^2\pi(i)}. \tag{3.19}$$

Note that, $\forall f$,

$$(f(i)-f(j))^2 = ((f(i)-\mathbb{E}_\pi[f]) - (f(j)-\mathbb{E}_\pi[f]))^2 = (\tilde{f}(i)-\tilde{f}(j))^2$$

where $\tilde{f}$ is a function with mean 0. Hence, in order to compare gap and gap_m we are allowed to consider just the denominators of (3.18) and (3.19), that is the terms $\sum_i \tilde{f}(i)^2\pi(i) = Var_\pi(f)$ and $\sum_i f(i)^2\pi(i) = \mathbb{E}_\pi[f^2]$.

Recall for brevity $\mathbb{E}_\pi[f] = \xi$. Then every f satisfying $f(k)=0$, $\forall k = m,...,n$, verifies

$$\frac{\mathbb{E}_\pi[f^2]}{\pi(f\neq 0)} = \mathbb{E}_\pi[f^2|f\neq 0] \geq (\mathbb{E}_\pi[f|f\neq 0])^2 = \frac{\xi^2}{\pi(f\neq 0)},$$

where we have used Holder's inequality. Then

$$\frac{1}{\mathbb{E}_\pi[f^2]} \leq \frac{\pi(f\neq 0)}{\xi^2} \leq \frac{1-\varepsilon}{\xi^2},$$

where the last inequality follows by the definition of m as $Q(1-\varepsilon)$. This gives

$$\frac{Var_\pi(f)}{\mathbb{E}_\pi[f^2]} = 1 - \frac{\xi^2}{\mathbb{E}_\pi[f^2]} \geq \varepsilon. \tag{3.20}$$

By the considerations made before on the ratio $\frac{\text{gap}}{\text{gap}_m}$ we conclude that

$$\text{gap}_m \geq \varepsilon \cdot \text{gap}.$$

■

3.1.4 Another result

In this section we are going to demonstrate just a technical Lemma that will be necessary for the proof of the main theorem.

Lemma 3.7. *Let (X_t) be a lazy irreducible Birth and death chain on Ω_n and suppose that there is an $\varepsilon \in (0, \frac{1}{16})$ such that*

$$t_{rel} < \varepsilon^4 \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]. \tag{3.21}$$

Then, for any fixed $\varepsilon \leq \alpha \leq \beta < 1-\varepsilon$,

$$\mathbb{E}_{Q(\alpha)}\left[\tau_{Q(\beta)}\right] \leq \frac{3}{2\varepsilon}\sqrt{t_{rel}\cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}. \tag{3.22}$$

Proof: Obviously it is sufficient to demonstrate (3.22) for $\alpha = \varepsilon$ and $\beta = 1 - \varepsilon$. Consider the random variable ν distributed according to the stationary distribution π restricted to the states in the set $[0, Q(\varepsilon)]$, that is

$$\nu(k) := \frac{\pi(k)}{\pi([0, Q(\varepsilon)])} \chi_{\{k \in Q(\varepsilon)\}}$$

and let w be the vector with components $w(k) := \frac{\chi_{\{k \in Q(\varepsilon)\}}}{\pi([0,Q(\varepsilon)])}$. By the reversibility of (X_t), we have for any state k

$$P^t(\nu, k) = \sum_i w(i) P^t(i,k)\pi(i) = \sum_i w(i) P^t(k,i)\pi(k) = (P^t w)(k) \cdot \pi(k).$$

Thus,

$$\begin{aligned} \|P^t(\nu, \cdot) - \pi\|_{TV} &= \frac{1}{2} \sum_{k=0}^{n} \pi(k) \left|(P^t w)(k) - 1\right| = \frac{1}{2} \|P^t(w - \bar{1})\|_{L^1(\pi)} \\ &\leq \frac{1}{2} \|P^t(w - \bar{1})\|_{L^2(\pi)} \end{aligned} \tag{3.23}$$

since we are in a finite space. As $\langle w - \bar{1}, \bar{1}\rangle_{L^2(\pi)} = 0$, the function $w - \bar{1}$ has to be decomposable as a combination of eigenfunctions with eigenvalues $\lambda_0, ..., \lambda_{n-1} \neq 1$ (because, as we said in the first chapter, $\bar{1}$ generates the space of the functions with eigenvalue 1). Calling λ the greatest of these eigenvalues (note that gap $= 1 - \lambda$), then, for all f orthogonal to $\bar{1}$, we have

$$\begin{aligned} \|P^t(f)\|_{L^2(\pi)} &= \Big(\sum_x \Big| \sum_y P^t(x,y) \Big(\sum_j \langle f, f_j \rangle_\pi f_j(y) \Big) \Big|^2 \pi(x) \Big)^{\frac{1}{2}} \\ &= \Big(\sum_x \Big| \sum_j \langle f, f_j \rangle_\pi \lambda_j^t f_j(x) \Big|^2 \pi(x) \Big)^{\frac{1}{2}} \\ &\leq \lambda^t \Big(\sum_x f^2(x) \pi(x) \Big)^{\frac{1}{2}} \end{aligned}$$

and thus

$$\|P^t(w - \bar{1})\|_{L^2(\pi)} \leq \lambda^t \|w - \bar{1}\|_{L^2(\pi)}.$$

Whence

$$\begin{aligned} \|P^t(\nu, \cdot) - \pi\|_{TV} \leq& \frac{1}{2} \lambda^t \|w - \bar{1}\|_{L^2(\pi)} \\ =& \frac{1}{2} \lambda^t \Big(\sum_{x \in [0, Q(\varepsilon)]} \Big(\frac{1}{\pi([0, Q(\varepsilon)])} - 1 \Big)^2 \pi(x) \\ &+ \sum_{x \notin [0, Q(\varepsilon)]} \pi(x) \Big)^{\frac{1}{2}} \\ \leq& \frac{1}{2} \lambda^t \sqrt{\frac{1}{\pi([0, Q(\varepsilon)])} - 1} \leq \frac{\lambda^t}{2\sqrt{\varepsilon}}. \end{aligned} \tag{3.24}$$

Define

$$t_\varepsilon := \left\lceil \frac{3}{2}\log\left(\frac{1}{\varepsilon}\right) t_{rel} \right\rceil$$

and observe that $\varepsilon \leq \frac{1}{16}$ implies

$$t_\varepsilon \leq 2\log\left(\frac{1}{\varepsilon}\right) t_{rel}.$$

Since $\log\left(\frac{1}{x}\right) \geq 1 - x$ for all $x \in (0,1]$, it follows that $\lambda^{t_\varepsilon} = \varepsilon^{\lceil \frac{3}{2}\log\frac{1}{\lambda} t_{rel}\rceil} \leq \varepsilon^{\frac{3}{2}}$ and so

$$\|P^{t_\varepsilon}(\nu,\cdot) - \pi\|_{TV} \leq \frac{\varepsilon}{2}. \tag{3.25}$$

On the other hand, calling $A := [Q(1-\varepsilon), n]$ and observing that $\pi(A) \geq \varepsilon$,

$$\|P^{t_\varepsilon}(\nu,\cdot) - \pi\|_{TV} \geq |P^{t_\varepsilon}(\nu, A) - \pi(A)| \geq \varepsilon - P^{t_\varepsilon}(\nu, A)$$

by the definition of total variation distance, and so

$$\mathbb{P}_\nu\left(\tau_{Q(1-\varepsilon)} \leq t_\varepsilon\right) \geq P^{t_\varepsilon}(\nu, A) \geq \varepsilon - \|P^{t_\varepsilon}(\nu,\cdot) - \pi\|_{TV} \geq \frac{\varepsilon}{2}. \tag{3.26}$$

Besides, being ν supported by $[0, Q(\varepsilon)]$, Chebychev inequality yields

$$\mathbb{P}_\nu\left(\tau_{Q(1-\varepsilon)} \leq t_\varepsilon\right) \leq \mathbb{P}_{Q(\varepsilon)}\left(\tau_{Q(1-\varepsilon)} \leq t_\varepsilon\right) \leq \frac{Var_{Q(\varepsilon)(\tau_{Q(1-\varepsilon)})}}{|\mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] - t_\varepsilon|^2}. \tag{3.27}$$

Combining (3.26) and (3.27),

$$\mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] \leq t_\varepsilon + \sqrt{\frac{2}{\varepsilon} Var_{Q(\varepsilon)}(\tau_{Q(1-\varepsilon)})}. \tag{3.28}$$

The variance under the square root is bounded above by $Var_0(\tau_{Q(1-\varepsilon)})$, since the $Q(1-\varepsilon)$-hitting time starting from 0 is the sum of the hitting time from 0 to $Q(\varepsilon)$ plus the hitting time from $Q(\varepsilon)$ to $Q(1-\varepsilon)$ and these two are independent. Thus Proposition 3.5 implies

$$\mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] \leq 2\log\left(\frac{1}{\varepsilon}\right) t_{rel} + \left(\frac{1}{\varepsilon}\right)\sqrt{2 t_{rel}\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]}. \tag{3.29}$$

The last effort is to rewrite the right hand side of this last equation in function of the only t_{rel} and $\tau_{Q(\frac{1}{2})}$. Using twice the hypothesis $t_{rel} < \varepsilon^4 \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]$ and remembering that $\varepsilon < \frac{1}{16}$ we have

$$\begin{aligned} \mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] &\leq (2\varepsilon^3 \log\left(\frac{1}{\varepsilon}\right) + \sqrt{2})\varepsilon \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right] \\ &\leq \frac{3}{2}\varepsilon \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]. \end{aligned} \tag{3.30}$$

The intuitive bound

$$\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right] \leq \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right] + \mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] \tag{3.31}$$

together with (3.30) shows that

$$\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right] \leq \frac{\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}{1-\frac{3}{2}\varepsilon}. \tag{3.32}$$

Plugging this back in (3.29) we obtain

$$\mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] \leq 2\log\left(\frac{1}{\varepsilon}\right)t_{rel} + \frac{1}{\varepsilon}\sqrt{\frac{2t_{rel}\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}{1-\frac{3}{2}\varepsilon}}.$$

Applying for the last time the hypothesis on t_{rel}, the fact that $\varepsilon < \frac{1}{16}$ and using the inequality (3.32) finally gives

$$\begin{aligned}\mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] &\leq \left(\frac{2\varepsilon^2\log\left(\frac{1}{\varepsilon}\right)+\frac{\sqrt{2}}{\varepsilon}}{\sqrt{1-\frac{3}{2}\varepsilon}}\right)\sqrt{t_{rel}\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]} \\ &\leq \frac{3}{2\varepsilon}\sqrt{t_{rel}\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]} \end{aligned} \tag{3.33}$$

as required. ■

3.2 A necessary and sufficient condition for the cut-off in B&D chains

3.2.1 A sufficient condition?

In section 2.1.4 we pointed out that a necessary condition for the existence of a cut-off for any sequence of Markov chains indexed by n with mixing times $t_{mix}^{(n)}$ is that

$$\lim_{n\to\infty} t_{mix}^{(n)} \cdot \mathrm{gap}^{(n)} \to \infty. \tag{3.34}$$

After having proved this fact, Yuval Peres conjectured in 2004 that for many natural classes of Markov chains this is also a *sufficient* condition. Nevertheless, many examples show that this is not true for any sequence of Markov chains: an important open problem is to characterize the classes of chains for which (3.34) implies the cut-off.

In 2006 Diaconis and Saloff-Coste ([7]) proved a variant of the conjecture in the case of continuous-time Birth and death chains. They verified that, when the convergence to equilibrium is measured in the so called **separation distance**, 'for continuous-time irreducible Birth and death chains, cut-off occurs iff (3.34) holds'; the separation distance between two distributions μ and ν is defined as

$$sep(\mu, \nu) := \sup_{x \in \Omega} \left(1 - \frac{\mu(x)}{\nu(x)}\right),$$

but it is not even an actual distance since it is not symmetric.

Following the article by Ding, Lubetzky and Peres [10] we are going to demonstrate that this is true also for arbitrary lazy irreducible discrete-time Birth and death chains with convergence to stationarity measured in total variation distance. This is implied by the following key theorem:

Theorem 3.8. *For any $0 < \varepsilon < \frac{1}{2}$ there exists an explicit $c_\varepsilon > 0$ such that every lazy irreducible discrete Birth and death chain satisfies*

$$t_{mix}(\varepsilon) - t_{mix}(1 - \varepsilon) \leq c_\varepsilon \sqrt{t_{rel} \cdot t_{mix}}. \tag{3.35}$$

In [10] it is also shown, with a bit more of work, that this result can be extended to the cases of δ-lazy chains ($r(x) > \delta$, $\forall x$) and continuous-time chains.

The proof of the theorem is put off to the next section. Now let's see how it implies what we were looking for.

Corollary 3.9. *Let $(X_t^{(n)})$ be a sequence of lazy irreducible discrete Birth and death chains. Then it exhibits a cut-off in total variation distance if and only if*

$$t_{mix}^{(n)} \cdot gap^{(n)} \xrightarrow{n \to \infty} \infty.$$

Proof (of the Corollary): Remembering the definition (2.1), (3.35) gives immediately the cut-off. In fact, for $0 < \varepsilon < \frac{1}{4}$,

$$\begin{aligned} 1 - \frac{t_{mix}^{(n)}(1-\varepsilon)}{t_{mix}^{(n)}(\varepsilon)} &\leq \frac{t_{mix}^{(n)}(\varepsilon) - t_{mix}^{(n)}(1-\varepsilon)}{t_{mix}^{(n)}} \\ &\leq c_\varepsilon \frac{\sqrt{t_{rel}^{(n)} \cdot t_{mix}^{(n)}}}{t_{mix}^{(n)}} \\ &= \frac{c_\varepsilon}{\sqrt{t_{mix}^{(n)} \cdot gap}} \xrightarrow{n \to \infty} 0. \end{aligned} \tag{3.36}$$

The fact that this condition is also necessary for cut-off is the result of Proposition 2.2. ∎

3.2.2 Proof of theorem 3.8

The hardest part of the theorem lies in the case where t_{rel} is much smaller than t_{mix}. Therefore we prove apart an intermediate result for this regime and this task will require all the general properties of the Birth and death chains obtained so far.

Theorem 3.10. *Let* (X_t) *be a lazy irreducible Birth and death chain such that*

$$t_{rel} < \varepsilon^5 \cdot t_{mix} \tag{3.37}$$

for some $0 < \varepsilon < \frac{1}{16}$. *Then*

$$t_{mix}(4\varepsilon) - t_{mix}(1-2\varepsilon) \leq \frac{6}{\varepsilon}\sqrt{t_{rel} \cdot t_{mix}}. \tag{3.38}$$

Proof: Recall the (not compromising) assumption

$$\mathbb{E}_n\left[\tau_{Q(\varepsilon)}\right] \leq \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right],$$

and define

$$t^- = t^-(\gamma) := \left\lfloor \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right] - \gamma\sqrt{t_{rel} \cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}\right\rfloor,$$
$$t^+ = t^+(\gamma) := \left\lceil \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right] + \gamma\sqrt{t_{rel} \cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}\right\rceil$$

(we will see that these are the extremes of the cut-off window).

First let's find a lower bound for $t_{mix}(1-2\varepsilon)$. Putting together hypothesis (3.37) with (3.12) gives

$$t_{rel} \leq 16\varepsilon^5 \cdot \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right] \leq \varepsilon^4 \cdot \mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]. \tag{3.39}$$

Now, Lemma 3.7 provides

$$\mathbb{E}_0\left[\tau_{Q(\varepsilon)}\right] \geq \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right] - \mathbb{E}_{Q(\varepsilon)}\left[\tau_{Q(1-\varepsilon)}\right] \geq \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right] - \frac{3}{2\varepsilon}\sqrt{t_{rel} \cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]},$$

while Proposition 3.5 guarantees that

$$Var_0(\tau_{Q(\varepsilon)}) \leq \frac{1}{1-\varepsilon} t_{rel} \cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right].$$

Therefore Chebychev inequality yields, $\forall \gamma > \frac{3}{2\varepsilon}$,

$$\begin{aligned}
\mathbb{P}_0\left(\tau_{Q(\varepsilon)} \leq t^-\right) &\leq \mathbb{P}_0\left(\left|\tau_{Q(\varepsilon)} - \mathbb{E}_0\left[\tau_{Q(\varepsilon)}\right]\right| \leq t^- - \mathbb{E}_0\left[\tau_{Q(\varepsilon)}\right]\right) \\
&\leq \frac{Var_0(\tau_{Q(\varepsilon)})}{\left(\frac{3}{2\varepsilon}\sqrt{t_{rel} \cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]} - \gamma\sqrt{t_{rel} \cdot \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}\right)} \\
&\leq \frac{2}{\left(\frac{3}{2\varepsilon} - \gamma\right)^2},
\end{aligned} \tag{3.40}$$

and this implies

$$\|P^{t^-}(0,\cdot)-\pi\|_{TV}\geq 1-\varepsilon-\mathbb{P}_0\left(\tau_{Q(\varepsilon)}\leq t^-\right)\geq 1-\varepsilon-\frac{2}{\left(\frac{3}{2\varepsilon}-\gamma\right)^2}.$$

Choosing $\gamma=\frac{2}{\varepsilon}$ (but it can be taken even smaller, since $\varepsilon<\frac{1}{16}$) we conclude that

$$t_{mix}(1-2\varepsilon)\geq \mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]-\frac{2}{\varepsilon}\sqrt{t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}. \tag{3.41}$$

The argument to bound from above $t_{mix}(4\varepsilon)$ is very similar. Taking again $0<\varepsilon<\frac{1}{16}$, Lemma 3.7 and Proposition 3.5 (in particular the bound (3.32)) show that

$$\mathbb{E}_n\left[\tau_{Q(\varepsilon)}\right]\leq\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]\leq\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]+\frac{3}{2\varepsilon}\sqrt{t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}, \tag{3.42}$$

$$Var_0(\tau_{Q(1-\varepsilon)})\leq\frac{1}{\varepsilon}t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]\leq\frac{t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}{\varepsilon\left(1-\frac{3}{2}\varepsilon\right)}\leq\frac{2}{\varepsilon}t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right],$$

$$Var_n(\tau_{Q(\varepsilon)})\leq\frac{1}{\varepsilon}t_{rel}\cdot\mathbb{E}_n\left[\tau_{Q(\varepsilon)}\right]\leq\frac{2}{\varepsilon}t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right].$$

Then, for $\gamma>\frac{3}{2\varepsilon}$ and for all $k\in\Omega$, we can use again Chebychev inequality in order to estimate $d(t^+)$:

$$\begin{aligned}\|P^{t^+}(k,\cdot)-\pi\|_{TV}&\leq 2\varepsilon+\mathbb{P}_0\left(\tau_{Q(1-\varepsilon)}>t^+\right)+\mathbb{P}_n\left(\tau_{Q(\varepsilon)}>t^+\right)\\&\leq 2\varepsilon+\frac{4}{\varepsilon\left(\gamma-\frac{3}{2\varepsilon}\right)^2}.\end{aligned} \tag{3.43}$$

Choosing $\gamma=\frac{35}{12\varepsilon}$ (with other room to spare)

$$t_{mix}(4\varepsilon)\leq\left\lceil\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]+\frac{35}{12\varepsilon}\sqrt{t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}\right\rceil.$$

We can assume $Q\left(\frac{1}{2}\right)>0$ (otherwise our estimates would give $\mathbb{E}_n\left[\tau_{Q(\varepsilon)}\right]=\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]=0$, then $Q(1-\varepsilon)=0$, $Q(\varepsilon)=n$ and so $\Omega=\{0\}$). Whence, since ε is small,

$$t_{mix}(4\varepsilon)\leq\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]+\frac{3}{\varepsilon}\sqrt{t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]}. \tag{3.44}$$

The last step is to rewrite the bounds in terms of the only t_{rel} and t_{mix}. For this purpose note that (3.32) gives

$$t_{rel}<\varepsilon^4\mathbb{E}_0\left[\tau_{Q(1-\varepsilon)}\right]\leq\frac{\varepsilon^4}{1-\frac{3}{2}\varepsilon}\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right],$$

and putting this into equation (3.41) itself leads to

$$t_{mix} \geq t_{mix}(1-2\varepsilon) \geq \left(1 - \frac{2\varepsilon}{\sqrt{1-\frac{3}{2}\varepsilon}}\right)\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right] \geq \frac{5}{6}\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]. \quad (3.45)$$

Melting together (3.41),(3.44) and (3.45) finally yields

$$t_{mix}(4\varepsilon) - t_{mix}(1-2\varepsilon) \leq \frac{5}{\varepsilon}\sqrt{t_{rel}\cdot\mathbb{E}_0\left[\tau_{Q(\frac{1}{2})}\right]} \leq \frac{6}{\varepsilon}\sqrt{t_{rel}\cdot t_{mix}}.$$

■

We are finally ready to prove (3.35). Remember that it says that for any $0 < \varepsilon < \frac{1}{2}$ we can find $c_\varepsilon > 0$ such that

$$t_{mix}(\varepsilon) - t_{mix}(1-\varepsilon) \leq c_\varepsilon\sqrt{t_{rel}\cdot t_{mix}}.$$

Proof (of Theorem 3.8): Let us first analyze the case of Theorem 3.10, that is when $t_{rel} < \varepsilon^5 \cdot t_{mix}$. Call $\varepsilon' := \frac{\varepsilon}{4}$; taking for example $\varepsilon' < \frac{1}{64}$, Theorem 3.10 gives

$$t_{mix}(\varepsilon)-t_{mix}(1-\varepsilon) \leq t_{mix}(4\varepsilon')-t_{mix}(1-2\varepsilon') \leq \frac{6}{\varepsilon'}\sqrt{t_{rel}\cdot t_{mix}} = \frac{24}{\varepsilon}\sqrt{t_{rel}\cdot t_{mix}}$$

so that (3.35) holds for $c_\varepsilon = \frac{24}{\varepsilon}$. But since the left hand side of (3.35) is monotone decreasing in ε by the definition of mixing time, this result can be extended to any value of $\varepsilon < \frac{1}{2}$ by choosing

$$c_1(\varepsilon) = 24\max\left\{\frac{1}{\varepsilon}, 64\right\}.$$

There is left to treat the regime $t_{rel} \geq \varepsilon^5 t_{mix}$. For $\varepsilon < \frac{1}{4}$, because of the submoltiplicativity of the mixing time (1.21), we have

$$\begin{aligned} t_{mix}(\varepsilon) - t_{mix}(1-\varepsilon) \leq t_{mix}(\varepsilon) &\leq \left\lceil \log_2 \epsilon^{-1} \right\rceil \sqrt{t_{mix}}\sqrt{t_{mix}} \\ &\leq \varepsilon^{-\frac{5}{2}} \log_2\left(\frac{1}{\varepsilon}\right)\sqrt{t_{mix}t_{rel}}. \end{aligned} \quad (3.46)$$

Again by monotonicity of the left hand side of (3.35) it is sufficient to take

$$c_2(\varepsilon) = \max\left\{\varepsilon^{-\frac{5}{2}}\log_2\left(\frac{1}{\varepsilon}\right), 64\right\}$$

(where 64 is obtained as $(1/4)^{-\frac{5}{2}}\log_2(\frac{1}{1/4})$) to extend the argument to all other epsilons.

In conclusion, setting

$$c_\varepsilon = \max\{c_1(\varepsilon), c_2(\varepsilon)\},$$

we obtain a constant valid in any case and for all $0 < \varepsilon < \frac{1}{2}$. ■

3.2.3 A tight cut-off window

Remember by (2.10) that a sequence ω_n is a cut-off window if it dominates the difference $t_{mix}(\varepsilon) - t_{mix}(1-\varepsilon)$ up to a constant and if $\omega_n = o(t_{mix})$. Therefore Theorem 3.8 implies that in the case of lazy irreducible Birth and death chains the geometric mean between $t_{mix}^{(n)}$ and $t_{rel}^{(n)}$ can be taken as cut-off window (the condition $\omega_n = o(t_{mix})$ is implied by the fact that $t_{mix} \cdot (t_{rel})^{-1} \to \infty$).

A natural question that arises is: can this result be improved? Does a smaller upper bound valid for all chains of this kind exist?
The answer is no. In fact it is possible to build explicitly examples of such chains where the window is exactly of order $\sqrt{t_{mix}^{(n)} \cdot t_{rel}^{(n)}}$. The construction is roughly the following: take any family $(X_t^{(n)})$ of lazy irreducible Birth and death chains with their mixing times $t_M^{(n)}$ and relaxation times $t_R^{(n)}$ (we know $t_R = o(t_M)$ by Corollary 3.9) which exhibits cut-off; it is possible to choose $\lambda_1, ..., \lambda_n \in [0,1)$ such that there exist new chains $(Y_t^{(n)})$ which have these numbers as non-trivial eigenvalues, which have mixing times satisfying $t_{mix}^{(n)} = (\frac{1}{2} + o(1))t_M^{(n)}$ and relaxation times $t_{rel}^{(n)} = \frac{1}{2}t_R^{(n)}$. This can be easily achieved by taking $(Y_t^{(n)})$ as a Birth and death chains having all death-probabilities equal to 0, an absorbing state at n and with non-trivial eigenvalues $\lambda_1, ..., \lambda_n$. It is possible to show that the sequence $(Y_t^{(n)})$ has cut-off window at least $\sqrt{t_M \cdot t_R}$. At this point it is sufficient to perturb a bit the transition probabilities in order to get irreducible chains and finally take their lazy versions: this way we double the values of $t_{rel}^{(n)}$ and $t_{mix}^{(n)}$ and obtain chains such that

$$\begin{cases} (1-\varepsilon)t_M^{(n)} \le t_{mix}^{(n)} \le (1+\varepsilon)t_M^{(n)} \\ |t_{rel}^{(n)} - t_R^{(n)}| \le \varepsilon \end{cases} \tag{3.47}$$

and with cut-off window of size $\sqrt{t_{mix}^{(n)} \cdot t_{rel}^{(n)}}$.

Chapter 4

Glauber dynamics for the Ising model

In this chapter our purpose is to prove the existence of the cut-off phenomenon for the Glauber dynamics on the Ising model with $\beta < 1$ when the underlying graph is complete, following the works by Levin, Luczak and Peres of December 2007, [16], and by Ding, Lubetzky and Peres of June 2008, [9]. The sequence of Markov chains will be indexed of course by n, the number of vertices of the graph.

4.1 The Ising model on the complete graph

4.1.1 The Curie-Weiss model

A **spin system** on a graph $G = (V, E)$ is a probability distribution on the state space $\Omega := \{-1, +1\}^V$. From a physical point of view, we can imagine that in each vertex of the graph there is a small magnet that can point upward (if the spin in this site is $+1$) or downward (if the spin is -1). Moreover if two vertices are connected by an edge, then the magnets on these vertices influence each other, trying to point in the same direction. This interaction is conditioned by a parameter $\beta \geq 0$ that can be physically interpreted as the inverse of the temperature of the environment: $\beta = \frac{1}{T}$. The lower is the temperature, the stronger is the interaction between the spins.

In the Ising model we associate to each possible configuration of spins $\sigma \in \Omega$ an **energy**:

$$H(\sigma) := -\sum_{\substack{v,w\in V:\\ v\sim w}} \sigma(v)\sigma(w),$$

where the sum is extended to all pair of vertices that are connected by an edge, and where with $\sigma(v)$ we indicate the spin of configuration σ in the site v. As one can see by the formula, the energy decreases as the number of pairs of neighbors whose spins agree grows.

Finally we assign a probability to each configuration σ:

$$\mu(\sigma) := \frac{e^{-\beta H(\sigma)}}{Z(\beta)}. \tag{4.1}$$

Here $Z(\beta)$ is the so called **partition function**, which ensures that μ is a probability measure:

$$Z(\beta) := \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)}. \tag{4.2}$$

The probability measure μ on Ω is know as the **Gibbs distribution** corresponding to the energy H.

We are going to deal just with a particular kind of graphs, that is the complete graphs. So, in our model every site will be connected with all the others by an edge. In particular, our energy will take into account all the possible $\binom{n}{2}$ pairs of vertices and will be normalized for convenience:

$$H(\sigma) := -\frac{1}{n} \sum_{(v,w)} \sigma(v)\sigma(w), \tag{4.3}$$

'(v, w)' meaning that the pairs of the kind (v_1, v_2) and (v_2, v_1) are counted just once.

Note that at infinite temperature ($\beta = 0$) the energy H plays no role; this means that the spins are completely independent between them and that the measure μ becomes uniform over all possible configurations in Ω. As β grows (the temperature decreases) H gets importance and μ favours configurations where more spins are aligned.

4.1.2 The Glauber dynamics

The process we want to consider is the **heat-bath Glauber dynamics** on the model we have just described. At each step we select a vertex $v \in V$ at random, we "delete" the current spin in v and replace it with a brand new spin chosen according to measure μ conditioned on the spins of all other vertices. So, we are changing only one spin in the whole system for each step. More precisely, define the **normalized magnetization** as

$$S(\sigma) := \frac{1}{n} \sum_{v \in V} \sigma(v); \tag{4.4}$$

then, if we have chosen vertex v to be updated, the probability of putting a "+"-spin in v is given by $p^+(S(\sigma) - \frac{\sigma(v)}{n})$, where

$$p^+(x) := \frac{e^{\beta x}}{e^{\beta x} + e^{-\beta x}} = \frac{1 + \tanh(\beta x)}{2}. \tag{4.5}$$

Analogously, the probability of updating v with a negative spin is given by $p^-(S(\sigma) - \frac{\sigma(v)}{n})$, with

$$p^-(x) := \frac{e^{-\beta x}}{e^{\beta x} + e^{-\beta x}} = \frac{1 - \tanh(\beta x)}{2}. \tag{4.6}$$

Since functions p^+ and p^- are always strictly positive, the chain is aperiodic and since it is possible to go from a configuration in Ω to any other configuration in at most n steps, the chain is also irreducible. Furthermore it is easy to check that the chain is reversible with respect to measure μ. We will denote the Glauber dynamics by $(X_t)_{t=0}^{\infty}$.

What happens if $\beta = 0$?

Once we have selected a vertex to be uploaded the probability of placing either a "+" or a "-" is the same. If we call the possible spins "0" and "1" instead of "+" and "-" and if we give an order to the vertices, it's not hard to recognize that we are dealing exactly with the lazy random walk on the n-dimensional hypercube we studied in Chapter 2! In this case we already know that there is a cut-off at time $\frac{1}{2} n \log n$ with window of size n.

4.1.3 Monotone coupling

The **grand coupling** is a coupling that involves several copies of our original chain. In particular, we let start a version X_t^σ of the Glauber dynamics from any possible configuration $\sigma \in \Omega$. The evolution of every copy is subject to a common source of randomness: at each step we choose with equal probability a vertex $v \in \{1, 2, ..., n\}$ and generate an uniform random variable $U \sim Unif([0, 1])$ independently from v; then $\forall \sigma \in \Omega$ we set

$$T^\sigma = \begin{cases} +1 & \text{if} \quad 0 \leq U \leq p^+\left(S(\sigma) - \frac{\sigma(v)}{n}\right) \\ -1 & \text{otherwise.} \end{cases} \tag{4.7}$$

and upload in v every instance of the chain with the respective spin generated this way.

This kind of coupling is also called **monotone coupling**. The reason for this name is that if we start two copies of the chain from configurations σ and $\tilde{\sigma}$ such that $\sigma(w) \leq \tilde{\sigma}(w)$, $\forall w \in V$, and if we update these two copies with the above rule, then $\forall t \geq 0$ we'll have $X_t^\sigma(w) \leq X_t^{\tilde{\sigma}}(w)$, $\forall w \in V$.

To lighten the notations, the bidimensional projection of the grand coupling on two generic coordinates starting by configurations σ and $\tilde{\sigma}$ will be indicated with $(X_t, \tilde{X}_t)$ instead of $(X_t^\sigma, \tilde{X}_t^{\tilde{\sigma}})$.

Finally we define the **Hamming distance** between two configurations σ and $\tilde{\sigma}$ as the number of sites in which they have different spins:

$$dist(\sigma, \tilde{\sigma}) := \frac{1}{2} \sum_{j=1}^{n} |\sigma(j) - \tilde{\sigma}(j)|. \tag{4.8}$$

Proposition 4.1. *The monotone coupling satisfies*

$$\mathbb{E}\left[dist(X_t,\tilde{X}_t)\right] \le \rho^t dist(\sigma,\tilde{\sigma}), \tag{4.9}$$

where

$$\rho := 1 - \frac{1}{n} + \tanh\left(\frac{\beta}{n}\right). \tag{4.10}$$

Proof: First of all we prove the thesis in the case $t = 1$ and $dist(\sigma,\tilde{\sigma}) = 1$. This means that σ and $\tilde{\sigma}$ differ only in one site, say in vertex v, with $\sigma(v) = -1$ and $\tilde{\sigma}(v) = +1$ without loss of generality. We start two copies of the Glauber dynamics $(X_t,\tilde{X}_t)$ and use the monotone coupling to update them. By definition, if v is selected we put there the same spin in both chains. If $w \neq v$ is selected (remember that $\sigma(w) = \tilde{\sigma}(w)$), we put a different spin in w if and only if

$$p^+(S(\sigma) - \sigma(w)) \le U \le p^+(S(\tilde{\sigma})) - \tilde{\sigma}(w)), \tag{4.11}$$

where $U \sim Unif([0,1])$.

Thus we can calculate the expectation of the distance between the two chains after one step:

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[dist(X_1,\tilde{X}_1)\right] = 1 - \frac{1}{n} + \sum_{w\neq v}\mathbb{P}\left(\{w \text{ is chosen and (4.11) holds}\}\right).$$

Note that $S(\tilde{\sigma}) - \frac{\tilde{\sigma}(w)}{n} = S(\sigma) - \frac{\sigma(w)}{n} + \frac{2}{n}$ if $w \neq v$. Setting $\frac{\hat{S}_w}{n} := S(\sigma) - \frac{\sigma(w)}{n}$ we have

$$\begin{aligned}\mathbb{P}\left(\{w \text{ is chosen and (4.11) holds}\}\right) &= \frac{1}{2n}\left[\tanh\left(\beta\frac{\hat{S}_w + 2}{n}\right) - \tanh\left(\beta\frac{\hat{S}_w}{n}\right)\right] \\ &\le \frac{1}{n}\tanh\frac{\beta}{n}.\end{aligned}$$

Therefore

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[dist(X_1,\tilde{X}_1)\right] = 1 - \frac{1}{n} + \tanh\left(\frac{\beta}{n}\right) = \rho \sim 1 - \frac{1}{n} + \frac{\beta}{n}. \tag{4.12}$$

Now take any two configurations σ and $\tilde{\sigma}$. Suppose that $dist(\sigma,\tilde{\sigma}) = k$, with $1 \le k \le n$. We know that there is a path $\sigma_0, \sigma_1, ..., \sigma_k$ such that $\sigma_0 = \sigma$, $\sigma_k = \tilde{\sigma}$ and $dist(\sigma_{i-1},\sigma_i) = 1$, $\forall i = 1,...,k$. Therefore

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[dist(X_1,\tilde{X}_1)\right] \le \sum_{i=1}^{k}\mathbb{E}\left[dist(X_1^{\sigma_i}, X_1^{\sigma_{i-1}})\right] \le \rho k = \rho\, dist(\sigma,\tilde{\sigma}). \tag{4.13}$$

Iterating (4.13) t-times we are done. ■

4.2 Related chains

4.2.1 Magnetization chain

We can associate to (X_t) another interesting and useful chain. We define the magnetization chain

$$S_t := S(X_t) = \frac{1}{n}\sum_{j=1}^{n} X_t(j). \tag{4.14}$$

It is a projection of the Glauber dynamics on the set

$$\Omega_S := \left\{-1, -1+\frac{2}{n}, ..., 1-\frac{2}{n}, 1\right\}.$$

(S_t) is again a Markov chain, since the transition probabilities of (X_t) depend actually on the only magnetization. In fact, the transition probabilities for the magnetization chain are given by

$$P_S(s,s') = \begin{cases} \dfrac{1+s}{2}p_-\left(s-\dfrac{1}{n}\right) & \text{if } s' = s - \frac{2}{n} \\ \dfrac{1-s}{2}p_+\left(s+\dfrac{1}{n}\right) & \text{if } s' = s + \frac{2}{n} \\ 1 - P_S\left(s, s-\frac{2}{n}\right) - P_S\left(s, s+\frac{2}{n}\right) & \text{if } s' = s. \end{cases} \tag{4.15}$$

Notice that $P_S(s,s') = P_S(-s,-s')$, so that the distribution of (S_t) starting from $s \in \Omega_S$ is exactly the same as that of $(-S_t)$ starting from $-s$: there is a strong symmetry around 0 in this process.

Using the same convention as before, we will call the magnetization of a chain started in $\tilde{\sigma}$ just $\tilde{S}_t$.

The magnetization chain will play a key role in the demonstration of the cut-off for the Glauber dynamics. In order to know better this useful process, we start studying how (S_t) is affected by the results obtained with the monotone coupling technique for (X_t).

Lemma 4.2. *For the monotone coupling $(X_t, \tilde{X}_t)$ on the original chain, we have*

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[|S_t - \tilde{S}_t|\right] \le \frac{2}{n}\rho^t dist(\sigma, \tilde{\sigma}), \tag{4.16}$$

where ρ is the same defined in (4.10).

Proof:

$$\begin{aligned} |S_t - \tilde{S}_t| &= \left|\frac{1}{n}\left(\sum_{j=1}^{n} X_t(j) - \tilde{X}_t(j)\right)\right| \\ &\le \frac{2}{n}\, dist(X_t, \tilde{X}_t). \end{aligned} \tag{4.17}$$

Taking the expectation and using the result of Proposition 4.1 gives (4.16). ■

Proposition 4.3. *$\forall s, \tilde{s} \in \Omega_S$ such that $s \geq \tilde{s}$, we have*

$$0 \leq \mathbb{E}_s [S_1] - \mathbb{E}_{\tilde{s}} [S_1] \leq \rho(s - \tilde{s}). \tag{4.18}$$

Furthermore, $\forall s, \tilde{s}$,

$$|\mathbb{E}_s [S_1] - \mathbb{E}_{\tilde{s}} [S_1] | \leq \rho|s - \tilde{s}|. \tag{4.19}$$

Proof: We can always see s and $\tilde{s}$ as the magnetizations of two configurations σ and $\tilde{\sigma}$ in Ω ($S(\sigma) = s$, $S(\tilde{\sigma}) = \tilde{s}$) such that $\sigma \geq \tilde{\sigma}$ (in the sense that $\sigma(v) \geq \tilde{\sigma}(v)$, $\forall v \in V$).

Starting a monotone coupling $(X_t, \tilde{X}_t)$ from these configurations we have

$$0 \leq \mathbb{E}_{\sigma,\tilde{\sigma}} \left[|S_1 - \tilde{S}_1| \right] \overset{\text{(monotonicity)}}{=} \mathbb{E}_\sigma [S_1] - \mathbb{E}_{\tilde{\sigma}}[\tilde{S}_1] \overset{(S_t \text{ is a MC})}{=} \mathbb{E}_s [S_1] - \mathbb{E}_{\tilde{s}}[\tilde{S}_1]$$

(note that this last term does not depend on the coupling). On the other hand, by Proposition 4.1,

$$\mathbb{E}_{\sigma,\tilde{\sigma}} \left[|S_1 - \tilde{S}_1| \right] = \mathbb{E}_{\sigma,\tilde{\sigma}} \left[\frac{2}{n} dist(X_1, \tilde{X}_1) \right] \leq \frac{2}{n} \rho \, dist(\sigma, \tilde{\sigma}) = \rho(s - \tilde{s}).$$

Putting together these two inequalities we obtain (4.18), while an analogous bound in the case $S(\sigma) \leq S(\tilde{\sigma})$ establishes (4.19). ■

4.2.2 Other results for the magnetization chain

Let's study in detail the drift of the magnetization chain (S_t).

Proposition 4.4. *For $\beta \leq 1$, $s \geq 0$, we have*

$$\mathbb{E} [S_{t+1} - S_t | S_t = s] \leq \frac{s(\beta - 1)}{n}. \tag{4.20}$$

Proof: Writing explicitly the value of the expectation we have

$$\begin{aligned} \mathbb{E} [S_{t+1} - S_t | S_t = s] &= \frac{2}{n} \left(\frac{1-s}{2} \right) p^+ \left(s + \frac{1}{n} \right) - \frac{2}{n} \left(\frac{1+s}{2} \right) p^- \left(s - \frac{1}{n} \right) \\ &= \frac{1}{n} \left[f_n(s) - s + \theta_n(s) \right], \end{aligned} \tag{4.21}$$

where

$$f_n(s) := \frac{1}{2} \left[\tanh \left(\beta \left(s + \frac{1}{n} \right) \right) + \tanh \left(\beta \left(s - \frac{1}{n} \right) \right) \right] \tag{4.22}$$

and

$$\theta_n(s) := -\frac{s}{2}\left[\tanh\left(\beta\left(s+\frac{1}{n}\right)\right) - \tanh\left(\beta\left(s-\frac{1}{n}\right)\right)\right]. \qquad (4.23)$$

Now, by the concavity of the function $\tanh(\cdot)$ we have that $f_n(s) \leq \tanh(\beta s)$ and, since it's an increasing function, $\theta_n(s)$ is negative $\forall s \geq 0$. Thus

$$\mathbb{E}\left[S_{t+1} - S_t | S_t = s\right] \leq \frac{1}{n}\left(\tanh(\beta s) - s\right) \leq \frac{s(\beta-1)}{n}.$$

■

From this proposition (and by a symmetry argument) we see that for $\beta < 1$ the magnetization tends always to decrease in absolute value; in fact it has a drift towards 0 that is stronger as we go far away from it.

Define

$$\tau_0 := \inf\left\{t \geq 0 : |S_t| \leq \frac{1}{n}\right\}. \qquad (4.24)$$

Clearly, if the number of vertices of the complete graph n is even, then $S_{\tau_0} = 0$, while if n is odd, then $S_{\tau_0} = \pm\frac{1}{n}$. The next lemma establishes that the probability of staying away from 0 decreases quite fast as time passes.

Lemma 4.5. *Suppose $\beta \leq 1$ and, for simplicity, n even. There is a constant $c > 0$ such that $\forall s \in \Omega_S$, $\forall u, t \geq 0$*

$$\mathbb{P}\left(|S_u| > 0, |S_{u+1}| > 0, ..., |S_{u+t}| > 0 \,|S_u = s\right) \leq \frac{cn|s|}{\sqrt{t}}. \qquad (4.25)$$

Proof: Because of the symmetry, it will be sufficient to demonstrate (4.25) for $s > 0$. By (4.20) we know that $\mathbb{E}\left[S_{t+1} - S_t | S_t\right] \leq 0$ as long as $S_t > 0$.

Looking at the transition probabilities we know that there exists a constant $b > 0$ such that $\mathbb{P}\left(S_{t+1} - S_t \neq 0 | S_t\right) \geq b$ for all times t and uniformly in n (this guarantees that the probability for S_t of not remaining still is always big enough). We can couple S_t with a symmetric random walk W_t on $\mathbb{Z}$ such that

- $W_0 = \frac{ns}{2}$;
- $\mathbb{P}\left(W_1 - W_0 \neq 0 | W_0 = w\right) = b > 0$ for all w;
- $\frac{nS_t}{2} \leq W_t$ till τ_0.

We just have to force S_t to go to the left (that is, to decrease of one unit) whenever W_t does. Applying Corollary 2.12 there exists a constant $c > 0$ such that

$$\begin{aligned}\mathbb{P}\left(S_u > 0, S_{u+1} > 0, ..., S_{u+t} > 0 \,|S_u = s\right) &\leq \mathbb{P}_{\frac{ns}{2}}\left(W_1 > 0, ..., W_t > 0\right) \\ &\leq \frac{cns}{\sqrt{t}}.\end{aligned}$$

■

4.2.3 Variance bound for the magnetization chain

Let's first state a general result.

Lemma 4.6. *Let Z_t be a Markov chain on $\mathbb{R}$; if there exists $\rho \in (0,1)$ such that for every $z, \tilde{z}$*

$$\left|\mathbb{E}_z\left[Z_t\right] - \mathbb{E}_{\tilde{z}}[\tilde{Z}_t]\right| \leq \rho^t |z - \tilde{z}|,$$

then $v_t := \sup_{z_0} Var_{z_0}(Z_t)$ verifies

$$v_t \leq v_1 \min\left\{t, \frac{1}{1-\rho^2}\right\}. \tag{4.26}$$

Proof: By the 'total variance formula' we have

$$Var_{z_0}(Z_t) = Var_{z_0}\left(\mathbb{E}_{z_0}\left[Z_t|Z_1\right]\right) + \mathbb{E}_{z_0}\left[Var_{z_0}(Z_t|Z_1)\right]. \tag{4.27}$$

Let Z_t and Z_t^* be two independent copies of the chain starting from z_0, and set $\varphi(z) = \mathbb{E}_z\left[Z_{t-1}\right]$. The first part of (4.27) can be bounded by

$$\begin{aligned}
Var_{z_0}\left(\mathbb{E}_{z_0}\left[Z_t|Z_1\right]\right) &= \frac{1}{2}\cdot 2\left(\mathbb{E}_{z_0}\left[\varphi^2(Z_1)\right] - \mathbb{E}^2_{z_0}\left[\varphi(Z_1)\right]\right)\\
&= \frac{1}{2}\{\mathbb{E}_{z_0}\left[\varphi^2(Z_1) - \varphi(Z_1)\varphi(Z_1^*)\right]\\
&\quad + \mathbb{E}_{z_0}\left[\varphi^2(Z_1^*) - \varphi(Z_1^*)\varphi(Z_1)\right]\}\\
&= \frac{1}{2}\mathbb{E}_{z_0}\left[(\varphi(Z_1) - \varphi(Z_1^*))^2\right]\\
&\overset{(hp)}{\leq} \frac{1}{2}\mathbb{E}_{z_0}\left[\rho^{2(t-1)}|Z_1 - Z_1^*|^2\right]\\
&\leq v_1\rho^{2(t-1)},
\end{aligned} \tag{4.28}$$

while we bound the second part of (4.27) with

$$\mathbb{E}_{z_0}\left[Var_{z_0}(Z_t|Z_1)\right] \leq \mathbb{E}_{z_0}\left[v_{t-1}\right] = v_{t-1}. \tag{4.29}$$

Putting all together yields

$$Var_{z_0}(Z_t) \leq v_1\rho^{2(t-1)} + v_{t-1}.$$

Iterating this inequality gives

$$v_t \leq v_1\rho^{2(t-1)} + v_{t-1} \leq \ldots \leq v_1 \sum_{j=0}^{t-1}\rho^{2j} \leq v_1 \min\left\{\frac{1}{1-\rho^2}, t\right\}.$$

■

Observe that we don't actually need *one* coupling valid for *all* pairs of states. It is enough that for every pair of states $z, \tilde{z}$ there exists a coupling such that

$$\mathbb{E}_{z,\tilde{z}}\left[\left|Z_1 - \tilde{Z}_1\right|\right] \leq \rho|z - \tilde{z}|. \tag{4.30}$$

Besides, if the state space of the Markov chain is discrete and if there is a path metric, it is sufficient that (4.30) holds for all pairs of neighbours.

Let's apply this lemma to our case:

Proposition 4.7. *If $\beta < 1$, then*

$$Var(S_t) = O\left(\frac{1}{n}\right) \tag{4.31}$$

as $n \to \infty$. If $\beta = 1$, then

$$Var(S_t) = O\left(\frac{t}{n^2}\right) \tag{4.32}$$

as $n \to \infty$.

Proof: Put together Lemma 4.3 and Lemma 4.6, and observe that

$$v_1 = \sup_{z_0} \mathbb{E}_{z_0}\left[(Z_1 - \mathbb{E}_{z_0}[Z_1])^2\right] \leq \sup_{z_0} \mathbb{E}_{z_0}\left[\left(\frac{4}{n}\right)^2\right] = \frac{16}{n^2}.$$

■

4.2.4 Expected spin value

In order to establish the cut-off at high temperature we also need to consider the number of positive and negative spins among subsets of the vertices.

Lemma 4.8. *Let $\beta < 1$. Then*

(i) $\forall \sigma \in \Omega$ and $\forall i = 1, 2, ..., n$, we have

$$|\mathbb{E}_\sigma[S_t]| \leq 2e^{-\frac{(1-\beta)t}{n}} \tag{4.33}$$

and

$$|\mathbb{E}_\sigma[X_t(i)]| \leq 2e^{-\frac{(1-\beta)t}{n}}. \tag{4.34}$$

(ii) $\forall A \subset V$ define

$$M_t(A) := \frac{1}{2}\sum_{j\in A} X_t(j); \tag{4.35}$$

then

$$|\mathbb{E}_\sigma[M_t(A)]| \leq |A|e^{-\frac{(1-\beta)t}{n}} \tag{4.36}$$

and, for some constant $c > 0$,

$$Var(M_t(A)) \leq cn. \tag{4.37}$$

(iii) $\forall \sigma \in \Omega$ *and* $\forall A \subset V$

$$\mathbb{E}_\sigma\left[|M_t(A)|\right] \leq n e^{-\frac{(1-\beta)t}{n}} + O(\sqrt{n}). \tag{4.38}$$

Proof: (i): Let $\bar{1}$ be the configuration with all "+"-spins and let $(X_t^+, \tilde{X}_t)$ be the monotone coupling with $X_0^+ = \bar{1}$ and $\tilde{X}_0 \sim \mu$. By Lemma 4.2 we have

$$\begin{aligned}
\mathbb{E}_{\bar{1}}\left[S_t^+\right] &\leq \mathbb{E}_{\bar{1},\mu}\left[|S_t^+ - \tilde{S}_t|\right] + \overbrace{\mathbb{E}_\mu\left[\tilde{S}_t\right]}^{=0} \\
&\leq \frac{2}{n}\rho^t \left(\sum_{\sigma\in\Omega} \left(\frac{1}{2} \sum_{j=1}^n \overbrace{|1-\sigma(j)|}^{\leq 2} \right) \mu(\sigma) \right) \\
&\leq 2\rho^t \leq 2e^{-\frac{(1-\beta)t}{n}}.
\end{aligned}$$

Then

$$\mathbb{E}_{\bar{1}}\left[S_t^+\right] = \frac{1}{n}\sum_{j=1}^n \mathbb{E}_{\bar{1}}\left[X_t^+(j)\right] = \mathbb{E}_{\bar{1}}\left[X_t^+(j)\right]$$

by symmetry, while by monotonicity we have

$$\mathbb{E}_\sigma\left[X_t(i)\right] \leq \mathbb{E}_{\bar{1}}\left[X_t^+(i)\right] \leq 2e^{-\frac{(1-\beta)t}{n}}.$$

Since the distribution of $-S_t$ started in $-s$ is the same as the distribution of S_t started in s, we also have

$$-2e^{-\frac{(1-\beta)t}{n}} \leq \mathbb{E}_\sigma\left[X_t(i)\right].$$

(ii): The first part follows directly from (i). For the second part, remember that the spins are positively correlated, so that

$$Var\Big(\sum_{i\in A} X_t(i)\Big) \leq Var\Big(\sum_{i=1}^n X_t(i)\Big) \leq n^2 Var(S_t) \leq cn,$$

by Proposition 4.7.

(iii): Let $(X_t, \tilde{X}_t)$ be the monotone coupling with $X_0 = \sigma$, $\tilde{X}_0 \sim \mu$. Then, with obvious notations,

$$\begin{aligned}
\mathbb{E}_\sigma\left[|M_t(A)|\right] &\leq \mathbb{E}_{\sigma,\mu}\left[|\tilde{M}_t(A) - M_t(A)|\right] + \mathbb{E}_\mu\left[|\tilde{M}_t(A)|\right] \\
&\leq \mathbb{E}_{\sigma,\mu}\left[dist(X_t, \tilde{X}_t)\right] + \sqrt{\mathbb{E}_\mu\left[\tilde{M}_t(A)^2\right]} \\
&\leq n\rho^t + \sqrt{\frac{n^2}{4}\mathbb{E}_\mu\left[\tilde{S}_t^2\right]} = n\rho^t + \frac{n}{2}\sqrt{Var_\mu(\tilde{S}_t)} \\
&= n\rho^t + O(\sqrt{n}) \\
&\leq ne^{-\frac{(1-\beta)t}{n}} + O(\sqrt{n}),
\end{aligned}$$

where for the second passage we have used Cauchy-Schwartz inequality, for the third we used Proposition 4.1 and the positive correlations among the spins $\{\tilde{X}_t(i)\}$, while for the estimate of the variance we used Proposition 4.7. ∎

4.2.5 Two coordinates chain

The last chain related to the Glauber dynamics we want to analyze is the **two coordinates chain**. Define

$$\Omega_0 := \left\{\sigma : |S(\sigma)| \leq \frac{1}{2}\right\} \tag{4.39}$$

and fix *once and for all* a configuration $\sigma_0 \in \Omega_0$. Let

$$\bar{u}_0 := |\{i : \sigma_0(i) = +1\}|, \quad \bar{v}_0 := |\{i : \sigma_0(i) = -1\}| \tag{4.40}$$

be the number of positive and negative spins in σ_0 respectively. Define also $\Lambda_0 := \{(u,v) : \frac{n}{4} \leq u, v \leq \frac{3n}{4}\}$ and observe that $\sigma_0 \in \Omega_0 \Leftrightarrow (\bar{u}_0, \bar{v}_0) \in \Lambda_0$. Given $\sigma \in \Omega$, we can define

$$U(\sigma) = U_{\sigma_0}(\sigma) := |\{i \in \{1,2,...,n\} : \sigma(i) = \sigma_0(i) = +1\}|, \tag{4.41}$$
$$V(\sigma) = V_{\sigma_0}(\sigma) := |\{i \in \{1,2,...,n\} : \sigma(i) = \sigma_0(i) = -1\}| \tag{4.42}$$

the number sites in which both σ and σ_0 have a "$+$"-spin (respectively, a "$-$"-spin).

Now, given a Glauber dynamics (X_t), we define the process

$$(U_t, V_t)_{t\geq 0} := (U(X_t), V(X_t))_{t\geq 0}. \tag{4.43}$$

Since we can write the magnetization of a configuration η as

$$S(\eta) = 2U(\eta) - 2V(\eta) - S(\sigma_0)$$

(note that $S(\sigma_0)$ here is just a constant) and since, as we have seen, the transition probabilities for (X_t) depend only on the magnetization of the current state, we can deduce that also $(U_t, V_t)_{t\geq 0}$ is a Markov chain. Its transition probabilities depend of course on σ_0 and its state space is $\{0,1,...,u_0\} \times \{0,1,...,v_0\}$. We will call its stationary measure π_2.

The connection between the original dynamics and this chain is shown in the following lemma:

Lemma 4.9.

$$\|\mathbb{P}_{\sigma_0}(X_t \in \cdot) - \mu\|_{TV} = \|\mathbb{P}_{(\bar{u}_0,\bar{v}_0)}((U_t, V_t) \in \cdot) - \pi_2\|_{TV}. \tag{4.44}$$

Proof: Let

$$\Omega_{(u,v)} := \{\sigma \in \Omega : (U(\sigma), V(\sigma)) = (u,v)\}.$$

Since $\mu(\,\cdot\,|\Omega_{(u,v)})$ and $\mathbb{P}_{\sigma_0}(X_t \in \cdot\,|(U_t,V_t)=(u,v))$ are uniform over $\Omega_{(u,v)}$, we have, for all $\eta \in \Omega$,

$$\mu(\eta) = \mu(\eta|\Omega_{(U(\eta),V(\eta))})\mu(\Omega_{(U(\eta),V(\eta))}) = \frac{\mu(\Omega_{(U(\eta),V(\eta))})}{|\Omega_{(U(\eta),V(\eta))}|};$$

therefore

$$\mathbb{P}_{\sigma_0}(X_t = \eta) - \mu(\eta) = \sum_{u,v} \frac{\chi_{\{\eta \in \Omega_{(u,v)}\}}}{|\Omega_{(u,v)}|}[\mathbb{P}_{\sigma_0}((U_t,V_t)=(u,v)) - \mu(\Omega_{(u,v)}].$$

Using the triangle inequality, summing over $\eta \in \Omega$ and changing the order of the summations

$$\|\mathbb{P}_{\sigma_0}(X_t \in \cdot) - \mu\|_{TV} \leq \|\mathbb{P}_{(\bar{u}_0,\bar{v}_0)}((U_t,V_t) \in \cdot) - \pi_2\|_{TV}.$$

For the reverse inequality note that

$$\pi_2((u,v)) = \mu(\{\sigma : U_{\sigma_0}(\sigma) = u, V_{\sigma_0}(\sigma) = v\});$$

if $B := \{\sigma : (U_{\sigma_0}(\sigma), V_{\sigma_0}(\sigma)) \in (\bar{U},\bar{V})\}$, where $(\bar{U},\bar{V}) \subset \{0,1,...,n\} \times \{0,1,...,n\}$, then

$$\begin{aligned}\|\mathbb{P}(X_t \in \cdot) - \mu\|_{TV} &\geq \sup_B |\mathbb{P}((X_t \in B) - \mu(B)| \\ &= \sup_{(\bar{U},\bar{V})} |\mathbb{P}\left((U_t,V_t) \in (\bar{U},\bar{V})\right) - \pi_2((\bar{U},\bar{V}))|.\end{aligned}$$

■

The main result we have to know about the two-coordinates chain is the following:

Proposition 4.10. *Let $\sigma, \tilde{\sigma} \in \Omega$ such that $S(\sigma) = S(\tilde{\sigma})$ and $R_0 := U(\tilde{\sigma}) - U(\sigma) > 0$. Let*

$$\Xi := \left\{\{\sigma : \min\{U(\sigma), \bar{u}_0 - U(\sigma), V(\sigma), \bar{v}_0 - V(\sigma)\} \geq \frac{n}{16}\right\}. \tag{4.45}$$

There exists a coupling $(X_t, \tilde{X}_t)$ of the Glauber dynamics with $X_0 = \sigma$, $\tilde{X}_0 = \tilde{\sigma}$, such that

(i) $S(X_t) = S(\tilde{X}_t)$, $\forall t \geq 0$;

(ii) let $R_t := U(\tilde{X}_t) - U(X_t)$; then $R_t \geq 0$ for all $t \geq 0$ and

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[R_{t+1} - R_t | X_t, \tilde{X}_t\right] \leq 0; \tag{4.46}$$

(iii) $\exists c > 0$ *independent from* n *such that, if* $X_t \in \Xi$ *and* $\tilde{X}_t \in \Xi$*, then*

$$\mathbb{P}_{\sigma,\tilde{\sigma}}\left(R_{t+1} - R_t \neq 0 | X_t, \tilde{X}_t\right) \geq c. \tag{4.47}$$

Proof: For any configuration $\sigma \in \Omega$ we can divide the vertices in four sets:

$$\begin{aligned} A(\sigma) &:= \{j \in \{1,2,...,n\} : \sigma_0(j) = +1,\ \sigma(j) = +1\} \\ B(\sigma) &:= \{j \in \{1,2,...,n\} : \sigma_0(j) = +1,\ \sigma(j) = -1\} \\ C(\sigma) &:= \{j \in \{1,2,...,n\} : \sigma_0(j) = -1,\ \sigma(j) = +1\} \\ D(\sigma) &:= \{j \in \{1,2,...,n\} : \sigma_0(j) = -1,\ \sigma(j) = -1\}. \end{aligned} \tag{4.48}$$

Clearly,

$$|A(\sigma)| = U(\sigma), \quad |B(\sigma)| = \bar{u}_0 - U(\sigma), \quad |C(\sigma)| = \bar{v}_0 - V(\sigma), \quad |D(\sigma)| = V(\sigma).$$

The description of the coupling is the following: X_t is updated as usual, that is a vertex v is chosen at random in V and the spin in v is replaced with a new spin

$$\zeta := \begin{cases} +1 & \text{with prob. } p^+\left(S_t - \frac{X_t(v)}{n}\right) \\ -1 & \text{with prob. } p^-\left(S_t - \frac{X_t(v)}{n}\right). \end{cases}$$

For $\tilde{X}_t$ we choose at random a vertex w such that $X_t(v) = \tilde{X}_t(w)$ and replace the spin in w with ζ. Furthermore, once $U(X_t) = U(\tilde{X}_t)$, we can force to take the vertices v and w from the same set among those in (4.48) ensuring $R_t \geq 0$ for all $t \geq 0$. Clearly condition (i) is automatically satisfied because of the construction of the coupling.

In order to study the behaviour of $R_{t+1} - R_t$ the following table is useful:

v	w	*spin*	$R_{t+1} - R_t$
$v \in B(X_t)$	$w \in D(\tilde{X}_t)$	$+1$	-1
$v \in C(X_t)$	$w \in A(\tilde{X}_t)$	-1	-1
$v \in A(X_t)$	$w \in C(\tilde{X}_t)$	-1	$+1$
$v \in D(X_t)$	$w \in B(\tilde{X}_t)$	$+1$	$+1$
All other cases			0

Now it is quite easy to describe the probabilities for the increments of R_t:

$$\mathbb{P}_{\sigma,\tilde{\sigma}}\left(R_{t+1} - R_t = -1 | X_t, \tilde{X}_t\right) = a(U_t, V_t, R_t), \tag{4.49}$$

$$\mathbb{P}_{\sigma,\tilde{\sigma}}\left(R_{t+1} - R_t = +1 | X_t, \tilde{X}_t\right) = b(U_t, V_t, R_t), \tag{4.50}$$

where

$$a(U_t, V_t, R_t) = \left(\frac{\bar{v}_0 - V_t}{n}\right)\left(\frac{U_t + R_t}{\bar{v}_0 + U_t - V_t}\right) p^- \left(S_t - \frac{1}{n}\right) + \\ + \left(\frac{\bar{u}_0 - U_t}{n}\right)\left(\frac{V_t + R_t}{\bar{u}_0 - U_t + V_t}\right) p^+ \left(S_t + \frac{1}{n}\right) \tag{4.51}$$

and

$$b(U_t, V_t, R_t) = \left(\frac{U_t}{n}\right)\left(\frac{\bar{v}_0 - V_t - R_t}{\bar{v}_0 + U_t - V_t}\right) p^- \left(S_t - \frac{1}{n}\right) + \\ + \left(\frac{V_t}{n}\right)\left(\frac{\bar{u}_0 - U_t - R_t}{\bar{u}_0 - U_t + V_t}\right) p^+ \left(S_t + \frac{1}{n}\right). \tag{4.52}$$

We obtain

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[R_{t+1} - R_t | X_t, \tilde{X}_t\right] = b(U_t, V_t, R_t) - a(U_t, V_t, R_t) \\ = -\frac{R_t}{n}\left[p^- \left(S_t - \frac{1}{n}\right) + p^+ \left(S_t + \frac{1}{n}\right)\right] \tag{4.53}$$

and in particular

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[R_{t+1} - R_t | X_t, \tilde{X}_t\right] \leq 0. \tag{4.54}$$

Finally, for point (iii), note that p^+ and p^- are uniformly distant by 0 and 1, so that there exists a constant $c > 0$ (uniform in n) for which

$$\mathbb{P}_{\sigma,\tilde{\sigma}}\left(R_{t+1} - R_t \neq 0 | X_t, \tilde{X}_t\right) \geq b(U_t, V_t, R_t) \geq c.$$

■

4.3 Cut-off for the Glauber dynamics

4.3.1 The main theorem

All the work done so far in this chapter will be used to demonstrate the existence of the cut-off for the Glauber dynamics. So, we have to consider the sequence of the complete graphs $G_n = (V_n, E_n)$, with $|V_n| = n$; on each graph we define the Ising model with its Gibbs' measure μ_n and the Glauber dynamics (X_t^n). For each of these chains we can define as usual the distance to the stationary measure as

$$d_n(t) := \max_{\sigma \in \Omega_n} \|\mathbb{P}_\sigma\left((X_t^n \in \cdot) - \mu_n\right)\|_{TV}$$

and the mixing time as

$$t_{mix}^n := \min\left\{t \geq 0 : d_n(t) \leq \frac{1}{4}\right\}.$$

For the ergodic theorem t_{mix}^n is finite for each fixed n, since $d_n(t) \to 0$ as $t \to \infty$. Nevertheless $d_n(t)$ will go to ∞ with n. Our aim is to understand the growth rate of the sequence t_{mix}^n.

Theorem 4.11. *Let $\beta < 1$. The Glauber dynamics for the Ising model on the n-complete graph has a cut-off at time $t_n := \frac{n \log n}{2(1-\beta)}$ with a window of size n.*

The proof of the theorem is rather long and will be split into two parts: the upper and the lower bound. They will be proved separately in the next sections.

Note again that the case where all the spins are independent (or equivalently $\beta = 0$) coincides with the random walk on the n-dimensional hypercube, and the result of the theorem agrees with the result of Theorem (2.13).

Yuval Peres conjectured a much general result. Our main theorem can be seen as a particular case of his hypothesis:

Conjecture 4.12. *Let (G_n) be a sequence of transitive graphs. If the Glauber dynamics on G_n has $t_{mix}^n = O(n \log n)$, then there is a cut-off.*

We didn't define the Glauber dynamics for a generic graph, but it's just the natural extension of our definition.

4.3.2 Upper bound

Theorem 4.13. *For $\beta < 1$*

$$\lim_{\gamma \to \infty} \limsup_{n \to \infty} d_n \left(\frac{n \log n}{2(1-\beta)} + \gamma n \right) = 0. \tag{4.55}$$

This upper bound is by all means the hardest part to demonstrate. The proof requires many "changes of strategy", so we will split it in three main phases and many sub-phases that will be briefly described in the following and deepened in the next sections.

Recall the useful definitions

$$\rho := 1 - \frac{1}{n} + \tanh\left(\frac{\beta}{n}\right),$$
$$\tau_0 := \min\left\{ t \geq 0 : |S_t| \leq \frac{1}{n} \right\},$$

and define also

$$t_n := \frac{1}{2(1-\beta)} n \log n, \tag{4.56}$$
$$t_n(\gamma) := t_n + \gamma n. \tag{4.57}$$

These are the main phases, summarized in Figure 4.1:

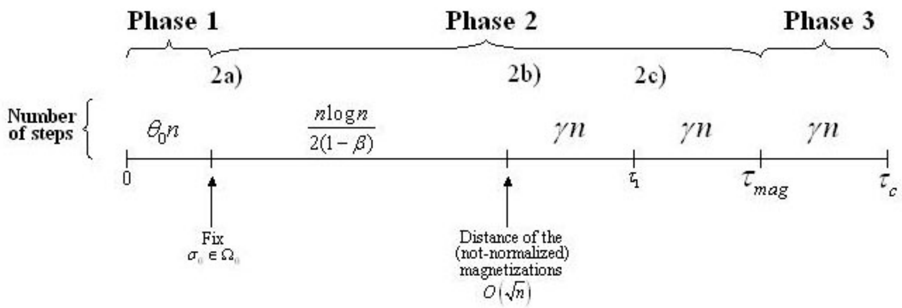

Figure 4.1: *Time line that describes the proof of Theorem 4.13. Note that these events are not certain, but happen with high probability.*

Phase 1: we will let the chain run for a *burn-in* period of $t_0 := \theta_0 n$ steps. After this period we will find ourselves in a "good" position (that is, in a configuration $\sigma_0 \in \Omega_0$ (remind its definition in (4.39))) with high probability:

$$d_n(t) \leq \max_{\sigma_0 \in \Omega_0} \|\mathbb{P}_{\sigma_0}(X_t \in \cdot) - \mu_n\|_{TV} + O\left(\frac{1}{n}\right).$$

Here we will fix once and for all a configuration $\sigma_0 \in \Omega_0$, with its $\bar{u}_0$ and $\bar{v}_0$, that will be fundamental in Phase 3. In the following we will ignore these t_0 steps, imaging to start in Ω_0 at time $t = 0$.

Phase 2: in this part we will start two instances of the chain, one starting by σ_0 and the other by any $\sigma \in \Omega$. The aim is to make the magnetizations of these two chains merge with high probability after $O\left(\frac{n \log n}{2(1-\beta)}\right)$ steps. We will use different kinds of coupling:

2a) for t_n steps we will use the monotone coupling and we will obtain the magnetizations to be quite close:

$$\mathbb{E}_{\sigma,\tilde{\sigma}}\left[\frac{n}{2}|S_{t_n} - \tilde{S}_t|\right] \leq cost.\sqrt{n}.$$

2b) Let $\tau_1 := \min\{t \geq t_n : \frac{n}{2}|S_t - \tilde{S}_t| \leq 1\}$. We will run the two chains independently till τ_1 and will show that

$$\mathbb{P}_{\sigma,\tilde{\sigma}}(\tau_1 > t_n + \gamma n) \leq O\left(\frac{1}{\sqrt{\gamma}}\right).$$

2c) If the number of "+"-spins in configuration X_{τ_1} will still be one more than those of configuration $\tilde{X}_{\tau_1}$, we will quickly (in γn steps) arrange the situation with a simple coupling.

Phase 3: Thanks to Lemma 4.9 it is possible to deal just with the two-coordinates chain. We will show that the hypothesis of Proposition 4.10 are verified in the time interval $[t_n(2\gamma), t_n(3\gamma)]$, so that we can start a coupling $(X_t, \tilde{X}_t)$ of the underlying Glauber dynamics and dominate $R_t := U(\tilde{X}_t) - U(X_t)$ with a simple random walk. This way, calling τ_c the first time the two chains meet, we will be able to show that

$$\mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(\tau_c > t_n(3\gamma)\right) \leq \frac{cost.}{\gamma} + O\left(\frac{1}{n}\right) \xrightarrow{\gamma,n\to\infty} 0.$$

This, together with the coupling corollary (Corollary 1.16), gives the theorem.

4.3.3 Phase 1

First of all let's demonstrate a general result.

Lemma 4.14. $\forall \Omega_0 \subset \Omega$,

$$\begin{aligned} d(t_0+t) &= \max_{\sigma\in\Omega} \|\mathbb{P}_\sigma\left(X_{t_0+t} \in \cdot\right) - \mu\|_{TV} \\ &\leq \max_{\sigma_0\in\Omega_0} \|\mathbb{P}_{\sigma_0}\left(X_t \in \cdot\right) - \mu\|_{TV} + \max_{\sigma\in\Omega} \mathbb{P}_\sigma\left(X_{t_0} \notin \Omega_0\right). \end{aligned} \tag{4.58}$$

Proof: $\forall A \subset \Omega$ we have

$$\begin{aligned} &\|\mathbb{P}_\sigma(X_{t_0+t} \in A - \mu(A))\|_{TV} \leq \\ &\leq \Big| \sum_{\sigma_0\in\Omega_0} \left[\mathbb{P}_\sigma\left(X_{t_0+t} \in A | X_{t_0} = \sigma_0\right) - \mu(A)\right] \mathbb{P}_\sigma\left(X_{t_0} = \sigma_0\right) \\ &+ \left[\mathbb{P}_\sigma\left(X_{t_0+t} \in A | X_{t_0} \notin \Omega_0\right) - \mu(A)\right] \mathbb{P}_\sigma\left(X_{t_0} \notin \Omega_0\right) \Big| \\ &\leq \sum_{\sigma_0\in\Omega_0} \left|\mathbb{P}_\sigma\left(X_{t_0+t} \in A | X_{t_0} = \sigma_0\right) - \mu(A)\right| \mathbb{P}_\sigma\left(X_{t_0} = \sigma_0\right) + \mathbb{P}_\sigma\left(X_{t_0} \notin \Omega_0\right). \end{aligned}$$

If we take the maximum over all possible $A \subset \Omega$

$$\begin{aligned} &\|\mathbb{P}_\sigma(X_{t_0+t} \in \cdot) - \mu\|_{TV} \\ &\leq \sum_{\sigma_0\in\Omega_0} \|\mathbb{P}_\sigma\left(X_{t_0+t} \in \cdot | X_{t_0} = \sigma_0\right) - \mu\|_{TV} \mathbb{P}_\sigma\left(X_{t_0} = \sigma_0\right) + \mathbb{P}_\sigma\left(X_{t_0} \notin \Omega_0\right) \\ &\leq \max_{\sigma_0\in\Omega_0} \|\mathbb{P}_{\sigma_0}\left(X_t \in \cdot\right) - \mu\|_{TV} + \mathbb{P}_\sigma\left(X_{t_0} \notin \Omega_0\right), \end{aligned}$$

where in the last step we used the Markov property. ■

Now we have just to apply the lemma with our $\Omega_0 = \{\sigma \in \Omega : |S(\sigma)| \leq \frac{1}{2}\}$. By Lemma 4.8 we know that there is $\theta_0 > 0$ such that

$$|\mathbb{E}_\sigma\left[S_{\theta_0 n}\right]| \leq 2e^{-(1-\beta)\frac{(\theta_0 n)}{n}} \leq \frac{1}{4}.$$

Thus, for n big enough,

$$\begin{aligned}\mathbb{P}_\sigma\left(X_{\theta_0 n}\notin\Omega_0\right)&=\mathbb{P}\left(|S_{\theta_0 n}|>\frac{1}{2}\right)\\&\le\mathbb{P}_\sigma\left(|S_{\theta_0 n}-\mathbb{E}_\sigma\left[S_{\theta_0 n}\right]|>\frac{1}{4}\right)\\&\le 16Var_\sigma(S_{\theta_0 n}).\end{aligned}$$

By Proposition 4.7 we finally have

$$d_n(\theta_0 n+t)\le\max_{\sigma_0\in\Omega_0}\|\mathbb{P}_{\sigma_0}\left(X_t\in\cdot\right)-\mu\|_{TV}+O\left(\frac{1}{n}\right). \tag{4.59}$$

After these t_0 steps we *fix* the configuration $\sigma_0:=X_{t_0}$ and its $\bar{u}_0,\bar{v}_0$ defined in (4.40). From now on, whenever we will mention σ_0, we will be talking about this precise configuration.

4.3.4 Phase 2

In Phase 2 we will use a particular coupling of the Glauber dynamics, letting two copies of the chain start one in σ_0 and the other in another generic $\sigma\in\Omega$; our aim will be to make the magnetizations of the two copies be the same after $O(t_n)$ steps.

Phase 2 is the part where the original chain has to spend more time. In fact, the heaviest part of the whole work is to force the magnetizations of the two copies to be sufficiently close, where sufficiently means that their difference (if we consider the not-normalized magnetizations) has to be at most of $O(\sqrt{n})$. To do that we will take advantage of the power of the drift towards 0 when the magnetization is far from it. This task itself will require $O(t_n)$ steps, that is also the cut-off order, so we can say that the heart of the problem is here.

For simplicity we will forget the first $t_0=\theta_0 n$ steps, assuming to start directly at time $t=0$ in the position $\sigma_0\in\Omega_0$. The following proposition is stated for any starting points $\sigma,\tilde{\sigma}$, but we can think σ as being our σ_0.

Proposition 4.15. *Let $\sigma,\tilde{\sigma}$ be any two configurations in Ω. There exist a coupling $(X_t,\tilde{X}_t)$ of the Glauber dynamics with $X_0=\sigma$, $\tilde{X}_0=\tilde{\sigma}$ and a constant $c>0$ (independent of $\sigma,\tilde{\sigma},n$) such that, if we define*

$$\tau_{mag}:=\min\{t\ge 0: S(X_t)=S(\tilde{X}_t)\}, \tag{4.60}$$

then

$$\mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_{mag}>t_n(2\gamma)\right)\le\frac{c}{\sqrt{\gamma}}. \tag{4.61}$$

Proof: The coupling will change its updating rules in the course of time. For the first t_n steps we will use the monotone coupling described in Section 4.1.3. Suppose, without loss of generality, $S(\sigma) > S(\tilde{\sigma})$. Define

$$\Delta_t := \frac{n}{2}|S_t - \tilde{S}_t| = \frac{1}{2}\Big|\sum_j X_t(j) - \tilde{X}_t(j)\Big| \in [0, n] \tag{4.62}$$

a measure of the different of the not-normalized magnetizations. By Lemma 4.2 there exists $c_1 > 0$ such that

$$\begin{aligned}
\mathbb{E}_{\sigma,\tilde{\sigma}}[\Delta_{t_n}] &= \frac{n}{2}\mathbb{E}_{\sigma,\tilde{\sigma}}\left[|S_t - \tilde{S}_t|\right] \le \rho^{t_n} dist(\sigma, \tilde{\sigma}) \\
&\le \left(1 - \frac{1-\beta}{n}\right)^{\frac{1}{2(1-\beta)} n \log n} n \\
&\le e^{-\left(\frac{1-\beta}{n}\frac{1}{2(1-\beta)} n \log n\right)} n \\
&\le c_1\sqrt{n}.
\end{aligned} \tag{4.63}$$

Now we change strategy: define

$$\tau_1 := \min\{t \ge t_n : |\Delta_t| \le 1\}. \tag{4.64}$$

As long as $t_n \le t < \tau_1$, we let X_t and $\tilde{X}_t$ evolve independently. By Lemma 4.3, since $S_t \ge \tilde{S}_t$ for all times till τ_1, $(S_t - \tilde{S}_t)_{t_n \le t \le \tau_1}$ has negative drift. Furthermore, because of the independence, the probability that $S_t - \tilde{S}_t \ne 0$ is uniformly bigger then 0. Therefore there is a simple random walk $(W_t)_{t \ge t_n}$ on $\mathbb{Z}$ that satisfies:

- $\mathbb{E}[W_{t+1} - W_t] = 0$,
- $W_{t+1} - W_t \le const.$,
- $n(S_{t_n} - \tilde{\sigma}_{t_n}) = W_{t_n}$,
- $n(S_t - \tilde{S}_t) \le W_t \qquad \forall t_n \le t < \tau_1$.

We can apply Corollary 2.12 and see that

$$\begin{aligned}
\mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_1 > t_n + \gamma' n | X_{t_n}, \tilde{X}_{t_n}\right) &\le \mathbb{P}_{\sigma,\tilde{\sigma}}\left(W_{t_n+1} > 0, ..., W_{t_n+\gamma' n} > 0 | X_{t_n}, \tilde{X}_{t_n}\right) \\
&\le \frac{n|S_{t_n} - \tilde{S}_{t_n}|}{\sqrt{\gamma' n}}.
\end{aligned}$$

Taking the expectation and plugging in (4.63) gives

$$\mathbb{P}_{\sigma,\tilde{\sigma}}(\tau_1 > t_n + \gamma' n) \le O\left(\frac{1}{\sqrt{\gamma'}}\right). \tag{4.65}$$

At this point the number of "+"-spins in X_{τ_1} can be equal to or one more than the number of "+"-spins in $\tilde{X}_{\tau_1}$. In the first case $\tau_1 = \tau_{mag}$, so we are done. If there is still one positive spin of difference, then we use a modified version of the monotone coupling: we put in a one-to-one correspondence the vertices with positive spin in $\tilde{X}_{\tau_1}$ with those with positive spin in X_{τ_1}, and pair all the other vertices arbitrarily. Then we let the system evolve with the rules of monotone coupling, updating together the matched vertices. We are allowed to use Lemma 4.2 replacing *dist*, the Hamming distance, with

$$dist' := \#\{\text{matched vertices with a different spin}\}$$

obtaining

$$\begin{aligned}\mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_{mag} > \tau_1 + \gamma'' n | X_{\tau_1}, \tilde{X}_{\tau_1}\right) &= \mathbb{P}_{\sigma,\tilde{\sigma}}\left(\Delta_{\tau_1+\gamma'' n} \geq 1 | X_{\tau_1}, \tilde{X}_{\tau_1}\right) \\ &\leq \mathbb{E}_{\sigma,\tilde{\sigma}}\left[\Delta_{\tau_1+\gamma'' n} | X_{\tau_1}, \tilde{X}_{\tau_1}\right] \\ &\leq \left(1 - \frac{1-\beta}{n}\right)^{\gamma'' n} \cdot 1 \\ &\leq e^{-(1-\beta)\gamma''}, \end{aligned} \tag{4.66}$$

where for the second line we have used the well known Markov inequality. In conlusion, taking $\gamma := \max\{\gamma', \gamma''\}$,

$$\begin{aligned}&\mathbb{P}_{\sigma,\tilde{\sigma}}(\tau_{mag} > t_n(2\gamma)) \\ &\quad = \mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_{mag} > t_n + 2\gamma n | \tau_1 > t_n + \gamma n\right) \mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_1 > t_n + \gamma n\right) \\ &\quad + \mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_{mag} > t_n + 2\gamma n | \tau_1 \leq t_n + \gamma n\right) \mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tau_1 \leq t_n + \gamma n\right) \\ &\quad \leq O\left(\frac{1}{\sqrt{\gamma}}\right) \cdot 1 + 1 \cdot e^{-(1-\beta)\gamma} \\ &\quad = O\left(\frac{1}{\sqrt{\gamma}}\right).\end{aligned}$$

■

4.3.5 Phase 3

Remember that Lemma 4.9, together with the fact that $\sigma_0 \in \Omega_0 \Leftrightarrow (\bar{u}_0, \bar{v}_0) \in \Lambda_0$, says

$$\max_{\sigma_0 \in \Omega_0} \|\mathbb{P}_{\sigma_0}(X_t \in \cdot) - \mu\|_{TV} = \max_{(\bar{u}_0,\bar{v}_0)\in\Lambda_0} \|\mathbb{P}_{(\bar{u}_0,\bar{v}_0)}((U_t, V_t) \in \cdot) - \pi_2\|_{TV}, \tag{4.67}$$

so that it will be sufficient to find a bound for the right hand side.

We would like to use Proposition 4.10 to be able to dominate the quantity $R_t := U(\tilde{X}_t) - U(X_t)$ with a simple random walk for γn steps. Let's show that all the hypothesis of the proposition are verified with high probability.

Define the events

$$H_1 := \{\tau_{mag} \le t_n(2\gamma)\} \tag{4.68}$$

$$H_2(t) := \{X_t \in \Xi, \tilde{X}_t \in \Xi\} \tag{4.69}$$

$$H_2 := \bigcap_{t \in I} H_2(t), \tag{4.70}$$

where I is the time interval $[t_n(2\gamma), t_n(3\gamma)]$.

Proposition 4.15 guarantees that H_1 is verified with probability at least $1 - \frac{c}{\sqrt{\gamma}}$; under this event we can take as starting configurations for the new coupling $\sigma = X_{t_n(2\gamma)}$ and $\tilde{\sigma} = \tilde{X}_{t_n(2\gamma)}$, which verify the first condition $S(\sigma) = S(\tilde{\sigma})$.

Without loss of generality we can assume $U(\sigma) - U(\tilde{\sigma}) > 0$.

There is only left to demonstrate that H_2 happens with high probability in order to get allowed to use point *(iii)* of the proposition for all times $t \in I$.

Lemma 4.16.

$$\mathbb{P}_{\sigma_0, \tilde{\sigma}}\left(H_2^c\right) = O\left(\frac{1}{n}\right). \tag{4.71}$$

Proof: Recall the definition

$$M_t(A) := \frac{1}{2} \sum_{j \in A} X_t(j)$$

and let

$$A_0 := \{i \in V : \sigma_0(i) = +1\}$$

$$B^* := \bigcup_{t \in I} \{|M_t(A_0)| \ge \frac{n}{32}\},$$

$$Y := \sum_{t \in I} \chi_{\{|M_t(A_0)| > n/64\}}$$

(clearly $|A_0| = \bar{u}_0$). $\tilde{M}_t(A)$, $\tilde{B}*$ and $\tilde{Y}$ have the very same definitions with $\tilde{X}_t$ instead of X_t.

If $M_t(A_0)$ goes once over the value $\frac{n}{32}$ in the interval I, it follows that it has to stay at least $\frac{n}{64}$ times over the value $\frac{n}{64}$, since its increments are in the set $\{-1, 0, +1\}$. Consequently $B^* \subset \{Y > \frac{n}{64}\}$ and

$$\mathbb{P}_{\sigma_0, \tilde{\sigma}}\left(B^*\right) \le \mathbb{P}_{\sigma_0, \tilde{\sigma}}\left(Y > \frac{n}{64}\right) \le \frac{c_0 \mathbb{E}_{\sigma_0, \tilde{\sigma}}\left[Y\right]}{n} \tag{4.72}$$

by Markov inequality. By Lemma 4.8 *(ii)*, we know that

$$\mathbb{P}_{\sigma_0, \tilde{\sigma}}\left(|M_t(A_0)| > \frac{n}{64}\right) = O\left(\frac{1}{n}\right),$$

so $\mathbb{E}[Y] = O(1)$, and equation (4.72) gives

$$\mathbb{P}_{\sigma_0,\tilde{\sigma}}(B^*) = O\left(\frac{1}{n}\right). \tag{4.73}$$

The same deductions lead to

$$\mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(\tilde{B}^*\right) = O\left(\frac{1}{n}\right). \tag{4.74}$$

Observe that

$$H_2^c(t) \subset \left\{|M_t(A_0)| \geq \frac{n}{16}\right\} \cup \left\{|\tilde{M}_t(A_0)| \geq \frac{n}{16}\right\}. \tag{4.75}$$

In fact if $U_t \leq \frac{n}{16}$, then $\bar{u}_0 - U_t \geq \frac{3n}{16}$ (since $\bar{u}_0 \geq \frac{n}{4}$ (because $\sigma_0 \in \Omega_0$)); thus

$$|M_t(A_0)| = \frac{1}{2}|U_t - (\bar{u}_0 - U_t)| \geq \frac{1}{2}((\bar{u}_0 - U_t) - U_t) \geq \frac{n}{16}.$$

Similarly if $\bar{u}_0 - U_t \leq \frac{n}{16}$, then $U_t \geq \frac{3n}{16}$ and thus $|M_t(A_0)| \geq \frac{n}{16}$ and the argument can be extended to V_t, $\bar{v}_0 - V_t$, $\tilde{U}_t$, $\bar{u}_0 - \tilde{U}_t$, $\tilde{V}_t$ and $\bar{v}_0 - \tilde{V}_t$.

Taking the union over the times in I, (4.75) implies, together with (4.74) and (4.73), that

$$\mathbb{P}_{\sigma,\tilde{\sigma}}(H_2^c) \leq \mathbb{P}_{\sigma,\tilde{\sigma}}(B^*) + \mathbb{P}_{\sigma,\tilde{\sigma}}\left(\tilde{B}^*\right) = O\left(\frac{1}{n}\right).$$

■

Recall that $R_t := U(\tilde{X}_t) - U(X_t)$. Thanks to Proposition 4.10 we know that, under the events H_1 and H_2, the process R_t can be dominated between $t_n(2\gamma)$ and $t_n(3\gamma)$ by a nearest-neighbor random walk until the first time when R_t visits 0.

Thus by Corollary 2.12, under H_1

$$\mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(\{\tau_c > t_n(3\gamma)\} \cap H_2 | X_{t_n(2\gamma)}, \tilde{X}_{t_n(2\gamma)}\right) \leq \frac{c_1|R_{t_n(2\gamma)}|}{\sqrt{n\gamma}}. \tag{4.76}$$

where τ_c is the first time t such that $U_t = \tilde{U}_t$. Taking the expectation gives

$$\mathbb{P}_{\sigma_0,\tilde{\sigma}}(\{\tau_c > t_n(3\gamma)\} \cap H_2 \cap H_1) \leq \frac{c_1\mathbb{E}_{\sigma_0,\tilde{\sigma}}\left[|R_{t_n(2\gamma)}|\right]}{\sqrt{n\gamma}}. \tag{4.77}$$

Writing $U_t = \frac{1}{2}(U_t - (\bar{u}_0 - U_t) + \bar{u}_0) = M_t(A_0) + \frac{\bar{u}_0}{2}$ and $\tilde{U}_t = \tilde{z}M_t(A_0) + \frac{\bar{u}_0}{2}$ we have $|R_t| \leq |M_t(A_0)| + |\tilde{M}_t(A_0)|$. Applying Lemma 4.8 *(iii)*,

$$\begin{aligned}
\mathbb{E}_{\sigma_0,\tilde{\sigma}}\left[|R_{t_n(2\gamma)}|\right] &\leq \mathbb{E}_{\sigma_0}\left[|M_{t_n(2\gamma)}(A_0)|\right] + \mathbb{E}_{\tilde{\sigma}}\left[|\tilde{M}_{t_n(2\gamma)}(A_0)|\right] \\
&\leq 2\left(ne^{-(1-\beta)\frac{t_n(2\gamma)}{n}} + O(\sqrt{n})\right) \\
&= O(\sqrt{n}).
\end{aligned} \tag{4.78}$$

Finally, using estimate (4.77),

$$\begin{aligned}\mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(\tau_c > t_n(3\gamma)\right) \leq & \mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(\{\tau_c > t_n(3\gamma)\} \cap H_1 \cap H_2\right) \\ & + \mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(H_1^c\right) + \mathbb{P}_{\sigma_0,\tilde{\sigma}}\left(H_2^c\right) \\ \leq & \frac{c_2}{\sqrt{\gamma}} + O\left(\frac{1}{n}\right). \end{aligned} \tag{4.79}$$

By equation (4.67) and by Corollary 1.16 we can conclude that

$$d(t_n + 3\gamma n) \leq \frac{c_2}{\sqrt{\gamma}} + O\left(\frac{1}{n}\right) \xrightarrow[\gamma\to\infty]{n\to\infty} 0, \tag{4.80}$$

that is the thesis of Theorem 4.13.

4.3.6 Lower bound

Theorem 4.17. *For $\beta < 1$*

$$\lim_{\gamma\to\infty} \liminf_{n\to\infty} d_n\left(\frac{n\log n}{2(1-\beta)} - \gamma n\right) = 1. \tag{4.81}$$

Proof: Since the magnetization chain is a projection of the Glauber dynamics (X_t), it will be sufficient to find a lower bound for the distance of (S_t) to its stationary distribution π_S.

Observing the definition (4.23), we see that $\theta_n(s) = O\left(\frac{1}{n}\right)$; thus, expanding $\tanh\left(\beta s + \frac{\beta}{n}\right)$ around βs in (4.22), we have, by (4.21),

$$\begin{aligned}\mathbb{E}\left[S_{t+1}|S_t = s\right] &\simeq s + \frac{1}{n}\left[\tanh(\beta s) + O\left(\frac{1}{n^2}\right) - s + O\left(\frac{1}{n^2}\right)\right] \\ &\simeq s + \frac{\beta s}{n} - \frac{\beta^3 s^3}{3n} - \frac{\beta s}{n} + O\left(\frac{1}{n^2}\right) \\ &\geq \rho s - \frac{s^3}{2n} + O\left(\frac{1}{n^2}\right).\end{aligned}$$

By symmetry it is also true that

$$\mathbb{E}\left[|S_{t+1}||S_t\right] \geq \rho|S_t| - \frac{|S_t|^3}{2n} + O\left(\frac{1}{n^2}\right). \tag{4.82}$$

Define

$$t^* := t_n - \frac{\alpha n}{(1-\beta)}. \tag{4.83}$$

We need the following lemma:

Lemma 4.18. *Take $S_0 = s_0 = s_0(\beta)$. If $s_0 < \frac{1-\beta}{3}$ and n is big enough, then*

$$\mathbb{E}_{s_0}\left[|S_{t^*}|\right] \geq B := \frac{s_0 e^\alpha}{2\sqrt{n}}. \tag{4.84}$$

Proof: Let $Z_t := |S_t|\rho^{-t}$, with $Z_0 = s_0$. Since $\rho^{-1} \leq 2$, by (4.82) it follows that, for n big enough,

$$\mathbb{E}_{s_0}\left[Z_{t+1}|Z_t\right] \geq Z_t - \rho^{-t}\frac{|S_t|^3 + O\left(\frac{1}{n}\right)}{n}$$

and therefore (remembering that $|S_t| \leq 1$)

$$\mathbb{E}_{s_0}\left[Z_t - Z_{t+1}|Z_t\right] \leq \rho^{-t}\frac{|S_t|^2 + O\left(\frac{1}{n}\right)}{n}. \tag{4.85}$$

By Lemma 4.8 we deduce, setting $A = V$, that

$$\mathbb{E}_{s_0}\left[|S_t|\right] \leq |s_0|\rho^t + c_1\frac{1}{\sqrt{n}}.$$

This and Proposition 4.7 give

$$\mathbb{E}_{s_0}\left[S_t^2\right] = \left(\mathbb{E}_{s_0}\left[S_t\right]\right)^2 + Var(S_t) \leq s_0^2\rho^{2t} + 2c_1\frac{|s_0|\rho^t}{\sqrt{n}} + \frac{c_3}{n}. \tag{4.86}$$

Taking the expectation in both sides of (4.85) and applying (4.86) yields

$$\mathbb{E}_{s_0}\left[Z_t - Z_{t+1}\right] \leq \frac{1}{n}\left[\rho^t s_0^2 + 2c_1\frac{|s_0|}{\sqrt{n}} + c_3\frac{\rho^{-t}}{n}\right] + O\left(\frac{1}{n^2}\right).$$

Adding the increments $\mathbb{E}_{s_0}\left[Z_k\right] - \mathbb{E}_{s_0}\left[Z_{k+1}\right]$ for $k = 0, 1, ..., t^* - 1$ we obtain

$$\begin{aligned} s_0 - \mathbb{E}_{s_0}\left[Z_{t^*}\right] &= \sum_{k=0}^{t^*}\mathbb{E}_{s_0}\left[Z_k - Z_{k+1}\right] \\ &\leq \frac{s_0^2}{n(1-\rho)} + \frac{2c_1|s_0|t^*}{n^{\frac{3}{2}}} + c_3\frac{\rho^{-t^*}}{n^2(1-\rho)} + O\left(\frac{t^*}{n^2}\right). \end{aligned}$$

Since $\rho^{-t^*} \leq \sqrt{n}$ we have

$$s_0 - \mathbb{E}_{s_0}\left[Z_{t^*}\right] \leq \frac{s_0^2}{1-\beta} + \frac{2c_2\log n}{\sqrt{n}} + \frac{c_4}{\sqrt{n}} \leq \frac{s_0}{2}$$

as long as $s_0 \leq \frac{1-\beta}{3}$ and n is big enough. Therefore

$$\mathbb{E}_{s_0}\left[|S_{t^*}|\right] \geq \frac{s_0\rho^{t^*}}{2} \geq \frac{s_0 e^{\alpha}}{2\sqrt{n}} = B.$$

■

By Proposition 4.7 we know that $\max\{Var_{s_0}(S_t), Var_{\mu}(S_t)\} \leq \frac{c_5}{n}$ and hence

$$\frac{B}{2} \leq \mathbb{E}_{s_0}\left[S_{t^*}\right] - \frac{s_0 e^{\alpha}}{4c_5}\sqrt{Var_{s_0}(S_{t^*})} \tag{4.87}$$

$$\frac{B}{2} \geq \mathbb{E}_{\mu}\left[S_t\right] + \frac{s_0 e^{\alpha}}{4c_5}\sqrt{Var_{\mu}(S_t)}. \tag{4.88}$$

Letting $D := \left[-\frac{B}{2}, \frac{B}{2}\right]$,

$$\begin{aligned}
\left\|\mathbb{P}_{s_0}\left(S_{t^*} \in \cdot\right) - \pi_S\right\|_{TV} \geq & \pi_S(D) - \mathbb{P}_{s_0}\left(|S_{t^*}| \in D\right) \\
\geq & \mathbb{P}_{\pi_S}\left(|S| \leq \mathbb{E}_\mu\left[S_t\right] + \frac{s_0 e^\alpha}{4c_5}\sqrt{Var_\mu(S_t)}\right) \\
& - \mathbb{P}_{s_0}\left(|S_{t^*}| \leq \mathbb{E}_{s_0}\left[S_{t^*}\right] - \frac{s_0 e^\alpha}{4c_5}\sqrt{Var_{s_0}(S_{t^*})}\right) \\
\geq & 1 - \frac{16c_5^2}{s_0^2 e^{2\alpha}} - \mathbb{P}_{s_0}\left(\left|\mathbb{E}_{s_0}\left[S_{t^*}\right] - S_{t^*}\right| \geq \frac{s_0 e^\alpha}{4c_5}\sqrt{Var_{s_0}(S_{t^*})}\right) \\
\geq & 1 - \frac{32c_5^2}{s_0^2 e^{2\alpha}}
\end{aligned} \tag{4.89}$$

where we have applied Chebychev inequality twice.

Since the last quantity in (4.89) tends to 1 as α goes to infinity, equation (4.81) is proved. ■

4.4 Near the critical point

4.4.1 Mixing time and cut-off in the critical case and in the low-temperature regime

Till now we have analyzed only the high temperature regime, that is the case $\beta < 1$, where the spins don't influence too much one another. What happens when we take β to be equal to 1 or greater? Is the order of the mixing time still $n \log n$? What happens to the cut-off point $\frac{n \log n}{2(1-\beta)}$? Does a cut-off still appear?

The articles [16] and [9] give again a complete answer to these questions.

For the critical case $\beta = 1$, the order of t_{mix} changes suddenly. In fact in the first article is shown that

Theorem 4.19. *If $\beta = 1$, then there are constants $c_1, c_2 > 0$ independent of n, such that*

$$c_1 n^{\frac{3}{2}} \leq t_{mix}^{(n)} \leq c_2 n^{\frac{3}{2}}. \tag{4.90}$$

The proof of this theorem uses the same techniques of the previous sections. In particular for the upper bound it is possible to show a coupling that makes the magnetizations of the two copies of the Glauber dynamics coalesce after $n^{\frac{3}{2}}$ steps; after that, it is possible to make agree the two configurations themselves in only other $O(n \log n)$ steps with another coupling. The second part of the proof shows a lower bound of order $n^{\frac{3}{2}}$ for the mixing time of the magnetization chain, which straight implies the same lower bound for the original dynamics.

The order of mixing time in the low-temperature regime, $\beta > 1$, was already known to be exponential in n (see e.g. [13]). The reason for this drastic slowing down is that the birth-and-death chain of the magnetization has no longer, after the critical β, the gaussian shape that will be described in the last chapter. In fact, as soon as β is greater than 1, two symmetric (with respect to 0) centers of mass appear and they drift further and further apart as the temperature decreases. The time to go from one center to the other is exponential and this implies the new order of t_{mix}.

Finally, in the second article ([9]), it is shown that in none of the two regimes $\beta = 1$ and $\beta > 1$ there is a cut-off. In the first instance the analysis of the spectral gap ensures that gap $= O(n^{\frac{3}{2}})$ and hence the non-existence of the cut-off follows from Proposition 2.2. Analogously in the low temperature case it results that gap $\cdot\, t_{mix} = O(1)$, and this again excludes the possibility of the cut-off.

4.4.2 Phase transition

Another natural question is how the phase transition between these states occurs around the critical value $\beta_c = 1$. Again [9] gives a satisfactory description of this phenomenon. Theorem 4.11 can in fact be refined this way:

Theorem 4.20. *Let $\delta = \delta(n) > 0$ be such that $\delta^2 n \xrightarrow{n\to\infty} \infty$. The Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 - \delta$ exhibits cut-off at time $\frac{n}{2\delta}\log(\delta^2 n)$ with window size $\frac{n}{\delta}$.*

Analogously the mixing time order of the critical point $\beta = 1$ can be extended to a little 'critical window':

Theorem 4.21. *Let $\delta = \delta(n)$ satisfy $\delta = O(\frac{1}{\sqrt{n}})$. The mixing time of the Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 \pm \delta$ has order $n^{\frac{3}{2}}$.*

Finally, for the supercritical regime we have

Theorem 4.22. *Let $\delta = \delta(n) > 0$ such that $\delta^2 n \xrightarrow{n\to\infty} \infty$. The mixing time of the Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 + \delta$ has order*

$$t_{\exp}(n) := \frac{n}{\delta}\left(\frac{n}{2}\int_0^{\zeta}\log\left(\frac{1+g(x)}{1-g(x)}\right)dx\right),$$

where $g(x) := \frac{\tanh(\beta x)-x}{1-x\tanh(\beta x)}$ and ζ is the only positive root of g. In particular, if $\delta \to 0$, the order of the mixing time is $\frac{n}{\delta}e^{(\frac{3}{4}+o(1))\delta^2 n}$, where the $o(1)$ tends to 0 as $n \to \infty$.

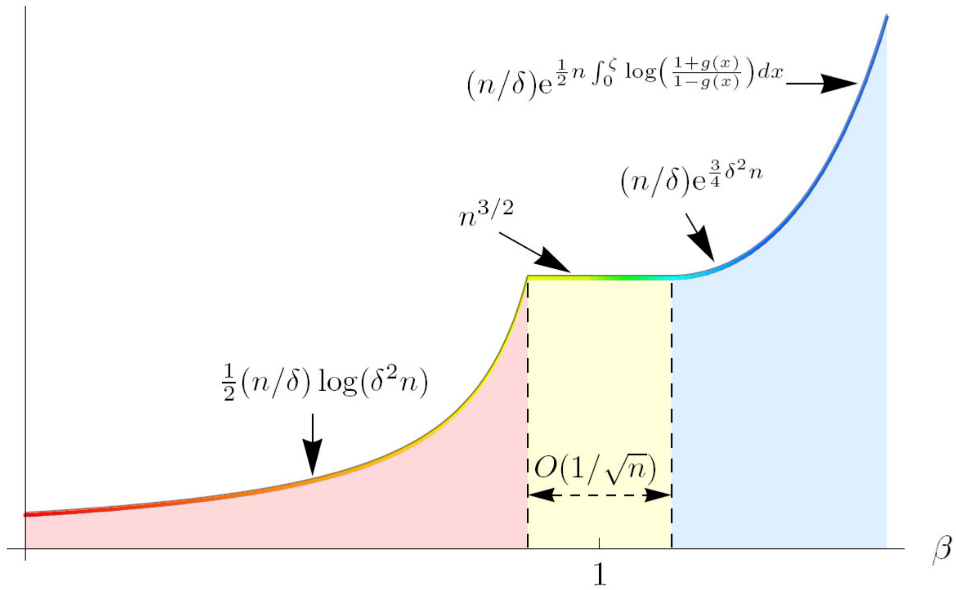

Figure 4.2: (from [9], pag.2) *Mixing time behaviour as a function of the inverse of the temperature* β*, with* n *fixed. Here* $\delta = |1-\beta|$ *and* ζ *is the unique positive roof of* $g(x) := \frac{\tanh(\beta x)-x}{1-x\tanh(\beta x)}$.

The meaning of these theorems is perfectly illustrated in Figure 4.2: fixing n, if the inverse of the temperature β is in an interval around 1 of order $\frac{1}{\sqrt{n}}$, then t_{mix} has order $n^{\frac{3}{2}}$, if β takes a value before this critical interval, then t_{mix} has order $n \log n$, while if β is after the interval, then t_{mix} assumes the exponential behaviour.

Chapter 5

An analytic attempt

The aim of this last chapter is to try to bound the spectral gap and the log-Sobolev constant (to be defined) of the continuous-time Glauber dynamics for the Ising model on the complete graph with a particular method. As we will see, this technique will turn to be completely successful in the first case but rather difficult to apply in the second.

5.1 Continuous-time chains

5.1.1 Discrete versus continuous time

A **continuous-time chain** $(X_t)_{t\geq 0}$ on a state space Ω is described by a matrix Q, called the **infinitesimal generator**, whose elements $q(x,y)$ are called **transition rates** and have the following properties:

(i) $-\infty < q(x,x) \leq 0$ for all $x \in \Omega$;

(ii) $q(x,y) \geq 0$ for all $x \neq y$;

(iii) $\sum_y q(x,y) = 0$ for all $x \in \Omega$.

Clearly $q(x) := -q(x,x) = \sum_{y\neq x} q(x,y)$; it can be thought as "the rate of leaving x".

The **heat kernel** H_t of the process is the object that specifies the actual probability of going from a state x to a state y after a time t: $H_t(x,y) = \mathbb{P}_x(X_t = y)$. It is defined as

$$H_t(x,y) := e^{tQ}. \tag{5.1}$$

We can give a very clear interpretation of the process. Define the $|\Omega| \times |\Omega|$ matrix J as: $J(x,y) = \frac{q(x,y)}{q(x)}$ for $y \neq x$ and $J(x,x) = 0$. Generate a path $x_0, x_1, ...$ of the discrete-time chain J. Then (X_t) visits this sequence of states remaining in each state x_i a time distributed as an exponential random variable of parameter $q(x_i)$.

Given a Markov chain P we can associate to it a Markov process (X_t), setting, for example, $Q = P - Id$. The random times between transitions are i.i.d. exponential random variables of unit rate, while the moves at these transition times are made according to P. In this case the heat kernel can be written as

$$H_t(x,y) = e^{t(P-Id)} = \sum_{k=0}^{\infty} \frac{e^{-t}t^k}{k!} P^k(x,y). \tag{5.2}$$

If up to time t we had k successes of exponentials of rate 1, the probability of staying in y starting in x is $P^k(x,y)$. Summing over all possible number of "jumps", that is exponential successes, we have the effective probability of going from x to y after a time t. The Markov property is transmitted from P to the new process (X_t) in the sense that, for $s \leq t$,

$$\mathbb{P}_x\left(X_t = y | \{\text{all the history up to time } s\}\right) = \mathbb{P}_x\left(X_t = y|X_s\right).$$

It is easy to see that if P has a stationary distribution π, then also (X_t) does. But one of the advantages of continuous-time chains is that they avoid the problem of periodicity. In fact, the Ergodic theorem 1.2 has the following equivalent:

Theorem 5.1. *Let P be an irreducible transition matrix, and let H_t be the corresponding heat kernel. Then there exists a unique probability distribution π such that $\pi H_t = \pi$ for all $t \geq 0$ and*

$$\max_{x\in\Omega} \|H_t(x,\cdot) - \pi\|_{TV} \xrightarrow{t\to\infty} 0.$$

5.1.2 Spectral gap and log-Sobolev constant for continuous time chains

Given a transition matrix P and the Markov process (X_t) associated to it with heat kernel H_t, we can define its Dirichlet form as

$$\mathcal{D}(f,f) = \sum_{x,y\in\Omega} (f(y) - f(x))^2 \pi(x) q(x,y).$$

If P is the reversible transition matrix that generates this process, and λ_2 is its greatest eigenvalue different from 1, we know from (1.29) that the spectral gap γ verifies

$$\gamma = 1 - \lambda_2 = \min_{\substack{f\in\mathbb{R}^\Omega \\ Var_\pi(f)\neq 0}} \left\{ \frac{\mathcal{D}(f,f)}{Var_\pi(f)} \right\}.$$

The utility of the spectral gap is shown by the following results, continuous-time versions of (1.31) and (1.32).

Lemma 5.2. *Let P be a reversible and irreducible transition matrix with spectral gap γ. Then for $f \in \mathbb{R}^\Omega$,*

$$\|H_t f - \mathbb{E}_\pi [f] \|_2^2 \leq e^{-2\gamma t} Var_\pi(f). \tag{5.3}$$

Lemma 5.3. *Let P be an irreducible transition matrix with spectral gap γ. Then, for the continuous-time chain associated to P, we have*

$$t_{mix}^{cont}\left(\frac{1}{2e}\right) \leq \frac{1}{2\gamma}\Big(2 + \log\Big(\frac{1}{\pi_{min}}\Big)\Big) \tag{5.4}$$

where $t_{mix}^{cont}(\cdot)$ is the obviously defined continuous equivalent of $t_{mix}(\cdot)$ and $\pi_{min} = \min_{x\in\Omega} \pi(x)$.

The definition of the **logarithmic Sobolev constant** α is similar to that of the spectral gap where the variance of f has been replaced by the entropy of f^2, where the **entropy** is

$$\mathrm{Ent}_\pi(f) = \mathbb{E}_\pi [f \log (f)] \quad \text{for} \quad f \geq 0,\, \pi(f) = 1. \tag{5.5}$$

Therefore

$$\alpha := \inf_{\substack{f\in\mathbb{R}^\Omega \\ \pi(f^2)=1}} \left\{ \frac{\mathcal{D}(f,f)}{\mathrm{Ent}_\pi(f^2)} \right\}. \tag{5.6}$$

The power of log-Sobolev constant is underlined by the upper bound of the next lemma, to be compared with the upper bound of (5.4).

Lemma 5.4. *Let P be a reversible and irreducible Markov chain. Then, for the continuous-time chain associated to P, we have*

$$\frac{1}{2\alpha} \leq t_{mix}^{cont}\left(\frac{1}{2e}\right) \leq \frac{1}{4\alpha}\left(4 + \log_+ \log \frac{1}{\pi_{min}}\right). \tag{5.7}$$

The proofs of these lemmas can be found, e.g., in [21].

5.1.3 Continuous time Glauber dynamics

Let's describe the continuous-time version of the Glauber dynamics for the Ising model on the complete graph examined in Chapter 4. We imagine to assign to each of the n sites a 'random clock' that rings at random times, distributed as independent exponential random variables of parameter 1. When the clock of site j rings, if the present configuration is σ, we try to update the spin in that place according to the Glauber dynamics, that is we put there a "+"-spin with probability $p^+(S(\sigma) - \frac{\sigma(j)}{n})$, where p_+ is defined in (4.5), and a "$-$"-spin else. This is equivalent to say that we have only a giant super-fast clock that rings every exponential time of parameter n, and when it rings we perform a step of the discrete-time Glauber dynamics. So the process is somehow accelerated, in the sense that in a unit of time we have tried on average to update all the spins.

Therefore the infinitesimal generator of the process is, for all $f \in \mathbb{R}^{\Omega}$,

$$\mathcal{L}f(\sigma) = \sum_{i} \sum_{x=\pm 1} \mathbb{P}\left(\sigma_i = x | \{\sigma_j\}_{j \neq i}\right) \left(f(\sigma^{i,x}) - f(\sigma)\right), \tag{5.8}$$

and its Dirichlet form

$$\mathcal{D}(f,f) := \sum_{\substack{\sigma \in \Omega \\ i=1,2,\ldots,n}} (f(\sigma^i) - f(\sigma))^2 \mathcal{L}(\sigma, \sigma^i)\pi(\sigma). \tag{5.9}$$

Let

$$\mu_i^{\sigma}(f) := \mu(f | \{\sigma_j\}_{j \neq i}) \tag{5.10}$$

be the expectation of f under the measure μ (the stationary distribution) once we have fixed all the spins except σ_i and analogously let

$$Var_i^{\sigma}(f) := Var(f | \{\sigma_j\}_{j \neq i}) \tag{5.11}$$

be the variance of f under the same measure.

Lemma 5.5.

$$\mathcal{D}(f,f) = \mu(f(-\mathcal{L}f)) = \sum_{i=1,2,\ldots,n} \mu(Var_i(f)) \tag{5.12}$$

Proof:

$$\mu(f(-\mathcal{L}f)) = \sum_{\sigma} \mu(\sigma) f(\sigma) \sum_{i} (f(\sigma) - \mu_i(f)) =$$
$$= \sum_{i} \left[\sum_{\sigma} \mu(\sigma) f^2(\sigma) - \sum_{\sigma} \mu(\sigma) f(\sigma) \mu_i(f) \right];$$

the second summatory in the square brackets is equal to

$$\mu(f\mu_i(f)) = \mu(\mu(f\mu_i(f) | \{\sigma_j\}_{j \neq i})) = \mu(\mu_i(f)\mu(f | \{\sigma_j\}_{j \neq i})) = \mu(\mu_i^2(f))$$

where we used very well know properties of the μ-expectation. Therefore

$$\mu(f(-\mathcal{L}f)) = \sum_{i} \left[\mu(f^2(\sigma) - \mu_i^2(f))\right] =$$
$$= \sum_{i} \mu\left(\mu(f^2(\sigma) - \mu_i^2(f)) | \{\sigma_j\}_{j \neq i}\right) = \sum_{i} \mu\left(Var_i(f)\right).$$

■

Corollary 5.6. *There exists $c > 0$, not depending on n, such that*

$$\sum_{i,\sigma} \mu(\sigma)(f(\sigma^i) - f(\sigma))^2 \le c\,\mathcal{D}(f,f) \tag{5.13}$$

Proof: First note that $Var_i^\sigma(f)$ (suppose $\sigma(i) = +$) is the variance of a Bernoulli variable that takes value $f(\sigma)$ with probability

$$p_i^\sigma = \mu(\sigma(i) = + | \{\sigma_j\}_{j\neq i})$$

and value $f(\sigma^i)$ with probability $1-p$. Therefore

$$Var_i^\sigma(f) = p_i^\sigma(1-p_i^\sigma)(f(\sigma^i) - f(\sigma))^2.$$

But for any σ and any i, we have

$$\begin{aligned} p_i^\sigma &= \frac{\mu(\sigma)}{\mu(\{\sigma_j\}_{j\neq i})} = \frac{\mu(\sigma)}{\mu(\sigma)+\mu(\sigma^i)} = \\ &= \frac{e^{\frac{\beta}{n}\sum_{j\neq i}\sigma_j}}{e^{\frac{\beta}{n}\sum_{j\neq i}\sigma_j} + e^{-\frac{\beta}{n}\sum_{j\neq i}\sigma_j}} = \frac{1}{1+e^{-2\frac{\beta}{n}\sum_{j\neq i}\sigma_j}} \\ &\ge \frac{1}{1+e^{2\beta}}, \end{aligned}$$

and analogously

$$1 - p_i^\sigma \ge \frac{1}{1+e^{2\beta}}.$$

Hence

$$\begin{aligned} \sum_{i,\sigma}\mu(\sigma)(f(\sigma^i)-f(\sigma))^2 &\le c\sum_{i,\sigma}\mu(\sigma)(f(\sigma^i)-f(\sigma))^2 p_i^\sigma(1-p_i^\sigma) \\ &= c\sum_{i,\sigma}\mu(\sigma)Var_i^\sigma(f) = c\sum_i \mu(Var_i(f)) \\ &= c\,\mathcal{D}(f,f). \end{aligned}$$

by Lemma (5.5). ■

5.1.4 The continuous-time magnetization chain

Remind that we can see the stationary distribution μ as a measure on the space $\Omega_S := \{-n, -n+2, ..., n-2, n\}$ defining

$$\mu(s) := \sum_{\sigma : S(\sigma)=s} \mu(\sigma).$$

We'd like to construct a continuous-time Birth and death process on Ω_s that has μ as stationary distribution. One possible choice of the rates of jump is the following: if the present state of the chain is $s > 0$, then the chain moves to the right with rate $b(s) := \frac{\mu(s+2)}{\mu(s)}$ and to the left with rate $d(s) := 1$; if $s < 0$, then the chain goes to the right with rate $b(s) := 1$ and to the left with rate $d(s) := \frac{\mu(s-2)}{\mu(s)}$; when in 0, the chain goes to the right or to the left both with rate 1. Obviously this chain is reversible for the measure μ. Observe that the chain has always a drift towards the central value 0 and that this drift becomes stronger as we go far away from 0.

Let's study the behaviour of $\mu(s)$. Remind that the number of configurations that have a certain magnetization s is $\binom{n}{\frac{n+s}{2}}$. For large values of n and small values of s and calling $m := \frac{s}{n}$ the normalized magnetization we have

$$\begin{aligned}\binom{n}{\frac{n+s}{2}} &\simeq \frac{\sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}}}{\sqrt{2\pi}\left(\frac{n+s}{2}\right)^{\frac{n+s}{2}+\frac{1}{2}}e^{-\frac{n+s}{2}}\sqrt{2\pi}\left(\frac{n-s}{2}\right)^{\frac{n-s}{2}+\frac{1}{2}}e^{-\frac{n-s}{2}}} = \\ &= \frac{n^{n+\frac{1}{2}}2^{n+1}}{\sqrt{2\pi}(n+s)^{\frac{n+s+1}{2}}(n-s)^{\frac{n-s+1}{2}}} = \\ &= \left(\frac{2^{n+1}}{\sqrt{n}\sqrt{2\pi}}\right)\frac{1}{(1+m)^{\frac{n+nm+1}{2}}(1-m)^{\frac{n-nm+1}{2}}} \simeq \\ &\simeq c(n)\left[\left(\frac{1-m}{1+m}\right)^{\frac{m}{2}}\frac{1}{(1-m^2)^{\frac{1}{2}}}\right]^n, \end{aligned} \tag{5.14}$$

where we have just used Stirling's approximation for the binomial coefficient. Therefore

$$\begin{aligned}\mu(s) &= \frac{\binom{n}{\frac{n+s}{2}}e^{\frac{\beta}{2n}s^2}}{Z} \\ &\simeq \frac{c(n)}{Z}e^{-n[-\beta\frac{m^2}{2}+\frac{1}{2}\log(1-m^2)-\frac{m}{2}\log\left(\frac{1-m}{1+m}\right)]} \simeq \\ &\simeq \frac{c(n)}{Z}e^{-n\left(\frac{1+\beta}{2}m^2\right)} = \frac{c(n)}{Z}e^{-\frac{1+\beta}{2n}s^2}.\end{aligned}$$

From this expression we can see that $\mu(s)$ has the shape of a "discrete gaussian measure" centered in 0 and with variance $\sim \frac{2n}{1+\beta}$.

Lemma 5.7. *For the magnetization chain, there exist a constant $\tilde{c} > 0$ not depending on n such that*

$$\frac{1}{\mathrm{gap}'} \leq \tilde{c}\,n, \tag{5.15}$$

where gap' *is the spectral gap of the process.*

Proof: For continuous time birth-and-death chains we have a powerful tool to estimate quite precisely the spectral gap: the Miclo's formulas. Define

$$B_+(i) := \sup_{x>i} \left(\sum_{y=i+1}^{x} \frac{1}{\mu(y)b(y)} \sum_{y \geq x} \mu(y) \right), \tag{5.16}$$

$$B_-(i) := \sup_{x<i} \left(\sum_{y=x}^{i-1} \frac{1}{\mu(y)b(y)} \sum_{y \leq x} \mu(y) \right), \tag{5.17}$$

and

$$B := \inf_{i \in \Omega_s} \left(B_+(i) \vee B_-(i) \right). \tag{5.18}$$

Then we know (see e.g. [3]) that the inverse of the spectral gap of the process is bounded by

$$\frac{B}{2} \leq \frac{1}{\text{gap}'} \leq 4B. \tag{5.19}$$

Let's start by evaluating $B_+(i)$ for a generical $i \geq 0$. Setting $k := 2/(1+\beta)$, for any $x > i$ we have

$$\begin{aligned}
\frac{Z}{c(n)} \sum_{y \geq x} \mu(y) &\simeq \int_x^\infty e^{-\frac{s^2}{kn}} ds \\
&= e^{-\frac{x^2}{kn}} \int_x^\infty e^{-\frac{(s-x)(s+x)}{kn}} ds \\
&\leq e^{-\frac{x^2}{kn}} \int_x^\infty e^{-\frac{2x(s-x)}{kn}} ds \\
&= e^{-\frac{x^2}{kn}} \frac{kn}{2x}
\end{aligned}$$

Then observe that for our dynamics

$$\begin{aligned}
\frac{Z}{c(n)} \sum_{y=i+1}^{x} \frac{1}{\mu(y)b(y)} &= \sum_{y=i+1}^{x} \frac{1}{\mu(y+1)} \\
&\simeq \int_i^x e^{\frac{s^2}{kn}} ds \\
&\leq e^{\frac{x^2}{kn}} (x-i).
\end{aligned}$$

So, for $i \geq 0$ (note that the constants simplify)

$$B_+(i) \leq \sup_{x>i} \left(e^{-\frac{x^2}{kn}} \frac{kn}{2x} \right) \left(e^{\frac{x^2}{kn}} (x-i) \right) = \sup_{x>i} \left(k \frac{x-i}{2x} \right) n. \tag{5.20}$$

Now analyze $B_-(i)$ for $i \leq 0$. Following the same steps as above we obtain

$$\frac{Z}{c(n)} \sum_{y \leq x} \mu(y) \leq -e^{-\frac{x^2}{kn}} \frac{kn}{2x}$$

and

$$\frac{Z}{c(n)}\sum_{y=x}^{i-1}\frac{1}{\mu(y)b(y)}=\frac{Z}{c(n)}\sum_{y=x}^{i-1}\frac{1}{\mu(y)}\leq e^{\frac{x^2}{kn}}(i-x).$$

So

$$B_-(i)\leq\sup_{x<i}\left(-k\frac{i-x}{2x}\right)n. \tag{5.21}$$

From equations (5.18),(5.21) and (5.20)

$$B\leq(B_+(0)\vee B_-(0))\leq\frac{k}{2}n \tag{5.22}$$

Using the bounds (5.19) and (5.22) we have that there exists $\tilde{c}$ independent from n such that

$$\frac{1}{\mathrm{gap}'}\leq\tilde{c}\,n.$$

■

Finally we exhibit the Dirichlet form for this process:

$$\mathcal{D}^{mag}(f,f)=\frac{1}{2}\sum_{s\in\Omega_s}(f(s+2)-f(s))^2[b(s)\mu(s)+d(s+2)\mu(s+2)].$$

Considering separately the cases $s<0$ and $s\geq 0$, writing explicitly the birth and death rates and putting all together again, we can write it in the shorter form

$$\mathcal{D}^{mag}(f,f)=\sum_{s\in\Omega_s}(f(s+2)-f(s))^2(\mu(s)\wedge\mu(s+2)). \tag{5.23}$$

5.2 Bound of the spectral gap

5.2.1 Conditioning on the magnetization

We want to bound the spectral gap of the Glauber dynamics on the n-complete graph in high temperature regime with a little trick: conditioning on a specific value of the magnetization, we will be able to bound the variance of any function on the space of the configurations and therefore show that the gap of the process doesn't depend on n, the size of the graph. Let's state this in a theorem.

Theorem 5.8. *For $\beta<1$, the spectral gap of the Glauber dynamics for the Ising model on the complete graph is bigger than a constant not depending on the number of the vertices of the graph.*

Using once more the 'total variance formula' we know that for any $s \in \{-n, -n+2/n, ..., n-2/n, n\}$ we can write the variance of any function f as

$$Var(f) = \mathbb{E}\left[Var(f|s)\right] + Var\left(\mathbb{E}\left[f|s\right]\right) \tag{5.24}$$

where with $Var(\cdot|s)$ and $\mathbb{E}[\cdot|s]$ we mean the variance and the expectation made on the configurations σ such that $S(\sigma) = s$.

Since the proof of the theorem is pretty long, we are going to bound separately the addends of (5.24) in the two next sections.

5.2.2 First bound via Bernoulli-Laplace model

The Bernoulli-Laplace model consists in n sites where we have to arrange r particles, or balls, with $r < n$. In each site there can be at most one ball. We assign to each particle a Poisson clock of rate 1 (this means that we have to wait an exponential time of rate 1 to have a "ring"). When the clock of i^{th}-particle rings, we choose at random one of the n sites and if that position is vacant we move the i^{th}-particle there. The stationary distribution π of this process is the uniform measure over all possible configurations of the r balls in the n sites. It is also well known (see e.g. [8]) that the spectral gap of the Bernoulli-Laplace model, $\text{gap}^{BL}_{n,r}$, is a constant not depending on n (in fact, its exact value is $\frac{1}{2}$!). Finally call

$$\mathcal{D}^{BL}_{n,r}(f,f) = \frac{1}{2n}\mathbb{E}_\pi\Big[\sum_{i,j}(f(\xi^{ij}) - f(\xi))^2\Big] \tag{5.25}$$

the Dirichlet form for this process (where we are taking the mean over the possible configurations ξ).

Lemma 5.9. *There exists a constant $k > 0$ not depending on n such that the first part of equation (5.24) is bounded by*

$$\mathbb{E}\left[Var(f|s)\right] \leq k \cdot \mathcal{D}(f,f). \tag{5.26}$$

Proof: If the magnetization of a certain configuration $\sigma \in \Omega$ is s, then we know that

$$n_+ := \frac{n+s}{2}, \quad n_- := \frac{n-s}{2}$$

are respectively the number of the "+" and "$-$"-spins of σ (clearly $n_+ + n_- = n$).

Note that, once we have fixed the magnetization, if we think to the positive spins as particles and the negative spins as holes, we can see every σ configuration as a Bernoulli-Laplace-model configuration with n sites and $r = n_+$ particles. Reminding that also the stationary distribution of the

Glauber dynamics on the complete graph is uniform over the states with a fixed magnetization, from the Poincarè inequality for the B-L model we have

$$Var(f|s) = Var^{BL}_{n,n_+}(f) \leq \frac{1}{(\text{gap}^{BL}_{n,n_+})}\mathcal{D}^{BL}_{n,n_+}(f,f)$$
$$\leq c\sum_i \frac{1}{2n}\sum_j \mathbb{E}_\pi\left[(f(\sigma^{ij}) - f(\sigma))^2|s\right] \tag{5.27}$$

where $\pi(\sigma) = 1/\binom{n}{n_+}$ for all σ with $S(\sigma) = s$. Note that, $\forall g : \Omega \to \mathbb{R}$,

$$\mathbb{E}\left[\mathbb{E}_\pi[g|s]\right] = \sum_{s\in\Omega_S} \mathbb{E}_\pi[g|s]\,\mathbb{P}(S = s)$$
$$= \sum_{s\in\Omega_S}\sum_{\substack{\sigma:\\ S(\sigma)=s}} \mathbb{P}(S = s)\frac{1}{\binom{n}{n_+}}g(\sigma)$$
$$= \sum_{\sigma\in\Omega} \mu(\sigma)g(\sigma) = \mu(g),$$

so that taking the expectation on both sides of (5.27) we obtain

$$\mathbb{E}\left[Var(f|s)\right] \leq c\frac{1}{2n}\sum_{i,j}\mu((f(\sigma^{ij}) - f(\sigma))^2)$$
$$= c\frac{1}{2n}\sum_{i,j}\mu((f(\sigma^{ij}) - f(\sigma^j) + f(\sigma^j) - f(\sigma))^2)$$
$$\leq \overbrace{c\frac{1}{n}\sum_{i,j}\mu((f(\sigma^{ij}) - f(\sigma^j))^2)}^{(A)} + \overbrace{c\frac{1}{n}\sum_{i,j}\mu((f(\sigma^j) - f(\sigma))^2)}^{(B)}. \tag{5.28}$$

where in the last inequality we used the fact that for any numbers a and b it is true that $(a + b)^2 \leq 2(a^2 + b^2)$.

In part (B) the factor $\frac{1}{n}$ and the summation over the i's delete each other, so that $(B) \leq c'\,\mathcal{D}(f,f)$ by Corollary 5.6.

For part (A) we need just a little bit more of work:

$$c\frac{1}{n}\sum_{\substack{\sigma\in\Omega\\ i,j}}\mu(\sigma)(f(\sigma^{ij}) - f(\sigma^j))^2 \overset{(\sigma^j=\eta)}{=} c\frac{1}{n}\sum_{i,j}\sum_{\eta\in\Omega}\frac{\mu(\eta^j)}{\mu(\eta)}\mu(\eta)(f(\eta^i) - f(\eta))^2$$
$$\leq \sup_{\substack{\sigma\in\Omega\\ x=1,\ldots,n}}\left(\frac{\mu(\sigma^x)}{\mu(\sigma)}\right) c\sum_i\sum_{\eta\in\Omega}\mu(\eta)(f(\eta^i) - f(\eta))^2 \leq c''\,\mathcal{D}(f,f)$$

by Corollary 5.6 and because the sup can be easily bounded with a constant not depending on n.

Putting all together yields

$$\mathbb{E}\left[Var(f|s)\right] \leq (c' + c'')\,\mathcal{D}(f,f).$$

■

5.2.3 Second bound via magnetization chain

Lemma 5.10. *There exists a constant $k' > 0$ not depending on n such that the second part of equation (5.24) is bounded by*

$$Var\left(\mathbb{E}\left[f|s\right]\right) \leq k' \cdot \mathcal{D}(f,f). \tag{5.29}$$

Proof: First of all note that $\mathbb{E}[f|s]$ is a function of the only magnetization. By Lemma 5.7 and equation (5.23) we can bound the variance of any function g of the magnetization with

$$Var(g(s)) \leq \frac{1}{\text{gap}'}\mathcal{D}^{mag}(g,g) \leq \tilde{c}\,n \sum_{s\in\Omega_s}(p(s)\wedge p(s+2))(g(s+2)-g(s))^2. \tag{5.30}$$

Of course we want to take $g(s) = \mathbb{E}[f|s]$, so that

$$Var(\mathbb{E}\left[f|s\right]) \leq \tilde{c}\,n \sum_{s\in\Omega_s}(p(s)\wedge p(s+2))(\mathbb{E}\left[f|s+2\right]-\mathbb{E}\left[f|s\right])^2. \tag{5.31}$$

Let's study the difference $\mathbb{E}[f|s+2] - \mathbb{E}[f|s]$. For $s > 0$ rewrite the first addend

$$\begin{aligned}
\mathbb{E}\left[f|s+2\right] &= \sum_{\substack{\sigma:\\ S(\sigma)=s+2}} \frac{\mu(\sigma)}{\mu(s+2)} f(\sigma) \overbrace{\left(\frac{\sum_i \chi_{\{\sigma_i=+1\}}}{\frac{s+n+2}{2}}\right)}^{(=1)} \\
&= \frac{1}{\frac{s+n+2}{2}} \sum_i \sum_{\substack{\sigma:\\ S(\sigma)=s+2}} \frac{\mu(\sigma)}{\mu(s+2)} f(\sigma)\chi_{\{\sigma_i=+1\}} \\
&\overset{(\eta=\sigma^i)}{=} \frac{1}{\frac{s+n+2}{2}} \sum_i \sum_{\substack{\eta:\\ S(\eta)=s}} \frac{\mu(\eta)}{\mu(s)}\frac{\mu(\eta^i)}{\mu(\eta)}\frac{\mu(s)}{\mu(s+2)} f(\eta^i)\chi_{\{\eta_i=-1\}}.
\end{aligned} \tag{5.32}$$

Writing $f(\eta^i) = [f(\eta^i) - f(\eta)] + f(\eta)$ we can split the above summation into two parts:

$$(I) = \frac{1}{\frac{s+n+2}{2}} \sum_i \sum_{\substack{\eta:\\ S(\eta)=s}} \frac{\mu(\eta)}{\mu(s)}\frac{\mu(\eta^i)}{\mu(\eta)}\frac{\mu(s)}{\mu(s+2)}[f(\eta^i) - f(\eta)]\chi_{\{\eta_i=-1\}}, \tag{5.33}$$

$$(II) = \frac{1}{\frac{s+n+2}{2}} \sum_i \sum_{\substack{\eta:\\ S(\eta)=s}} \frac{\mu(\eta)}{\mu(s)}\frac{\mu(\eta^i)}{\mu(\eta)}\frac{\mu(s)}{\mu(s+2)} f(\eta)\chi_{\{\eta_i=-1\}}. \tag{5.34}$$

Since

$$\frac{\mu(\eta^i)}{\mu(\eta)} =: K(s)$$

depends only on the magnetization s, and since

$$\sum_i \chi_{\{\eta_i=-1\}} = \frac{n-s}{2}$$

for the η's with magnetization s, we have that part (II) is equal to

$$(II) = \left(\frac{\mu(s)}{\mu(s+2)} K(s) \frac{n-s}{n+s+2}\right) \sum_{\substack{\eta: \\ S(\eta)=s}} \frac{\mu(\eta)}{\mu(s)} f(\eta) = c(s,n)\mathbb{E}\left[f|s\right];$$

now, taking the test function $f \equiv 1$, we have that

$$\mathbb{E}\left[f|s+2\right] = 1 = (I) + (II) = 0 + c(s,n)\mathbb{E}\left[f|s\right] = c(s,n),$$

and hence $c(s,n) \equiv 1$. It follows that

$$\mathbb{E}\left[f|s+2\right] - \mathbb{E}\left[f|s\right] = (I).$$

Looking at (I) we understand that we are just doing the expectation of $[f(\eta^i) - f(\eta)]$ with a particular probability distribution; in fact

$$\sum_i \sum_{\substack{\eta\in\Omega: \\ S(\eta)=s}} \frac{\mu(\eta^i)}{\mu(s+2)} \frac{\chi_{\{\eta_i=-1\}}}{\frac{s+n+2}{2}} = \sum_{\substack{\sigma\in\Omega: \\ S(\sigma)=s+2}} \frac{\mu(\sigma)}{\mu(s+2)} \overbrace{\left(\sum_i \frac{\chi_{\{\sigma_i=+1\}}}{\frac{s+n+2}{2}}\right)}^{(=1)} = 1.$$

Because of this fact we can use Schwartz inequality to bound our difference $(\mathbb{E}\left[f|s+2\right] - \mathbb{E}\left[f|s\right])^2 = (I)^2$:

$$\begin{aligned}(I)^2 &\overset{(Sch.)}{\leq} \frac{1}{\frac{s+n+2}{2}} \sum_i \sum_{\substack{\eta: \\ S(\eta)=s}} \frac{\mu(\eta)}{\mu(s)} \overbrace{\frac{\mu(\eta^i)}{\mu(\eta)} \frac{\mu(s)}{\mu(s+2)}}^{(\leq cost.)} \chi_{\{\eta_i=-1\}} [f(\eta^i) - f(\eta)]^2 \\ &\leq \frac{2\,c}{s+n+2} \mathbb{E}\Big[\sum_i (\triangle_i f)^2 \Big| s\Big] \qquad (5.35)\end{aligned}$$

where $\triangle_i f(\sigma) = f(\sigma^i) - f(\sigma)$.

For $s < 0$ we can do the very same kind of calculations but rewriting this time $\mathbb{E}\left[f|s\right]$ in function of $\mathbb{E}\left[f|s+2\right]$ (changing a "$-$"-spin in a "$+$"-spin) to obtain

$$(\mathbb{E}\left[f|s+2\right] - \mathbb{E}\left[f|s\right])^2 \leq \frac{2\,c'}{n-s} \mathbb{E}\left[\sum_i (\triangle_i f)^2 \Big| s\right]. \qquad (5.36)$$

Finally, putting together (5.31), (5.35) and (5.36), we have

$$\begin{aligned} Var(\mathbb{E}[f|s]) &\le 2c\tilde{c}\sum_{s\ge 0}\mu(s)\frac{n}{s+n+2}\mathbb{E}\left[\sum_i(\triangle_i f)^2\Big|s\right]+ \\ &+2c'\tilde{c}\sum_{s<0}\mu(s+2)\frac{n}{n-s}\mathbb{E}\left[\sum_i(\triangle_i f)^2\Big|s+2\right] \\ &\le k'\,\mathbb{E}\left[\sum_i(f(\sigma^i)-f(\sigma))^2\right] \\ &\le k'\,\mathcal{D}(f,f) \end{aligned} \tag{5.37}$$

where last inequality is by Corollary 5.6 again. ■

The last thing to do is to put equations (5.26) and (5.37) in equation (5.24). In conclusion we can find a constant K independent from n such that for any function f on the state space Ω we have

$$Var(f) \le K\mathcal{D}(f,f). \tag{5.38}$$

Therefore, by the variational representation (1.29), the spectral gap of the Glauber dynamics for the Ising model on the complete graph with n vertices and $\beta < 1$ is bigger than a constant which doesn't depend on the size of the graph, proving Theorem 5.8.

5.3 Bound of the log-Sobolev constant

The result of the last section is not actually a novelty. The fact that the spectral gap of the Glauber dynamics for the Ising model on the complete graph is a constant was already know, but it served us as a testing ground for the conditioning-on-the-magnetization method. The really interesting thing to do is to try to bound the log-Sobolev constant with the same method.

5.3.1 Conditioning on the magnetization, again

Recall the definition of the log-Sobolev for the Glauber dynamics:

$$\alpha := \inf_{\substack{f\in\mathbb{R}^\Omega \\ \mu(f^2)=1}}\left\{\frac{\mathcal{D}(f,f)}{\mathrm{Ent}_\mu(f^2)}\right\}.$$

where the entropy is, for any positive function with $\mu(f)=1$,

$$\mathrm{Ent}_\mu(f) = \sum_{\sigma\in\Omega} f(\sigma)\log(f(\sigma))\mu(\sigma). \tag{5.39}$$

In order to bound α, we are going to use the very same method of the last sections, that is conditioning on a specific value of the magnetization in order to divide the entropy of any function $f \geq 0$, $\mu(f) = 1$, as the sum of two parts, and estimate these parts separately. Write

$$\begin{aligned}\mathrm{Ent}_\mu(f) &= \mathbb{E}\left[\mathbb{E}_\mu\left[f\log(f)|s\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}_\mu\left[\frac{f}{\mathbb{E}_\mu\left[f|s\right]}\mathbb{E}_\mu\left[f|s\right]\log\left(\frac{f}{\mathbb{E}_\mu\left[f|s\right]}\mathbb{E}_\mu\left[f|s\right]\right)\Big|s\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}_\mu\left[f|s\right]\cdot\mathrm{Ent}_\mu\left(\frac{f}{\mathbb{E}_\mu\left[f|s\right]}\Big|s\right)\right] + \mathrm{Ent}_\pi(\mathbb{E}_\mu\left[f|s\right]). \qquad (5.40)\end{aligned}$$

5.3.2 Bound via magnetization chain, again

First we prove the equivalent of Lemma 5.7 for the log-Sobolev constant of the magnetization chain.

Lemma 5.11. *For the magnetization chain, there exists a constant $c > 0$ not depending on n such that*

$$\frac{1}{\alpha_S} \leq c\,n, \qquad (5.41)$$

where α_S is the log-Sobolev constant of the process.

Proof: Once again we use Miclo's formulas, this time to estimate α'. Define

$$A_+(i) := \sup_{x>i}\Big(\sum_{y=i+1}^{x}\frac{1}{\mu(y)b(y)}\Big)\log\Big(\frac{1}{\sum_{y\geq x}\mu(y)}\Big)\sum_{y\geq x}\mu(y), \qquad (5.42)$$

$$A_-(i) := \sup_{x<i}\Big(\sum_{y=x}^{i-1}\frac{1}{\mu(y)b(y)}\Big)\log\Big(\frac{1}{\sum_{y\leq x}\mu(y)}\Big)\sum_{y\leq x}\mu(y)), \qquad (5.43)$$

and

$$A := \inf_{i\in\Omega_s}\left(A_+(i)\vee A_-(i)\right). \qquad (5.44)$$

Then we know (see e.g. [3]) that the inverse of the log-Sobolev constant of the process is bounded by

$$\frac{A}{20} \leq \frac{1}{\alpha_S} \leq 20A. \qquad (5.45)$$

Remember that $\mu(s)$ has the shape of a "discrete gaussian measure" centered in 0 and with variance $\sim \frac{2n}{1+\beta}$. Since we are going to take $i = 0$, by symmetry it will be sufficient to bound $A_+(0)$. Let's bound the factors of

$A_+(0)$ separately (from now, c will be a constant not depending on n that will take, if necessary, different values along the proof):

$$\sum_{y\geq x}\mu(y)\simeq\int_x^n\frac{c}{\sqrt{n}}e^{-\frac{s^2}{kn}}ds\leq c\frac{\sqrt{n}}{x}e^{-\frac{x^2}{kn}},$$

as in Lemma 5.7. Then we improve a bit the second bound:

$$\begin{aligned}\sum_{y=i+1}^{x}\frac{1}{\mu(y)b(y)}&\simeq c\sqrt{n}\int_i^x e^{\frac{s^2}{kn}}ds=c\sqrt{n}e^{\frac{x^2}{n}}\int_i^x e^{-\frac{(x-s)(x+s)}{kn}}ds\\&\leq c\sqrt{n}e^{\frac{x^2}{n}}\int_i^x e^{-\frac{(x+i)}{kn}(x-s)}ds\\&=c\sqrt{n}e^{\frac{x^2}{n}}\int_0^{x-i} e^{-\frac{(x+i)}{kn}y}ds\\&=c\frac{n\sqrt{n}}{x+i}e^{\frac{x^2}{kn}}\Big(1-e^{-\frac{x^2-i^2}{kn}}\Big).\end{aligned}$$

To bound the logarithm we consider separately two cases: when $x=o(\sqrt{n})$, $\sum_{y\geq x}\mu(y)$ is greater than the mass of a gaussian after a standard deviation (which is greater than a constant), so that

$$\log\Big(\frac{1}{\sum_{y\geq x}\mu(y)}\Big)\leq cost. \tag{5.46}$$

When $x\geq c\sqrt{n}$, we have

$$\begin{aligned}\sum_{y\geq x}\mu(y)&\simeq\int_x^n\frac{c}{\sqrt{n}}e^{-\frac{s^2}{kn}}ds\overset{(y=\frac{s}{\sqrt{n}})}{=}c\int_{\frac{x}{\sqrt{n}}}^{\sqrt{n}}e^{-\frac{y^2}{k}}dy\\&\geq c\int_{\frac{x}{\sqrt{n}}}^{2\frac{x}{\sqrt{n}}}e^{-\frac{y^2}{k}}dy\geq c\frac{x}{\sqrt{n}}e^{-4\frac{x^2}{kn}}\\&\geq c\,e^{-4\frac{x^2}{kn}},\end{aligned}$$

and thus

$$\log\Big(\frac{1}{\sum_{y\geq x}\mu(y)}\Big)\leq c\frac{x^2}{n}. \tag{5.47}$$

Putting all together we obtain, in the case $x=o(n)$,

$$\begin{aligned}A_+(0)&\leq\sup_{x>0}c\cdot\frac{n^2}{x^2}\Big(1-e^{-\frac{x^2}{kn}}\Big)\\&\simeq\sup_{x>0}c\cdot\frac{n^2}{x^2}\Big(1-\Big(1-\frac{x^2}{kn}+o\Big(\frac{x^2}{n}\Big)\Big)\\&\leq c\,n\end{aligned}$$

by Taylor's formula. On the other hand, if $x \geq c\,n$,

$$A_+(0) \leq \sup_{x>0} c\,\frac{n^2}{x^2}\left(1 - e^{-\frac{x^2}{kn}}\right)\frac{x^2}{n} \leq cn.$$

Hence $A \leq cost.$, and by (5.45) we finally have the desired bound. ■

Lemma 5.12. *There exists a constant $k > 0$ not depending on n such that the second part of equation (5.40) is bounded by*

$$\mathrm{Ent}\left(\mathbb{E}_\mu\left[f|s\right]\right) \leq k \cdot \mathcal{D}(f,f). \tag{5.48}$$

Proof: The proof is exactly the same as that of Lemma 5.10. By Lemma 5.11 the log-Sobolev constant α_S for the magnetization chain verifies

$$\frac{1}{\alpha_S} \leq c\,n, \tag{5.49}$$

so that, for any function g on Ω_s, we have

$$\mathrm{Ent}(g(s)) \leq \frac{1}{\alpha_S}\mathcal{D}^{mag}(g,g) \leq c\,n \sum_{s\in\Omega_s} (p(s) \wedge p(s+2))(g(s+2) - g(s))^2.$$

From this point to the end, taking $g(s) = \mathbb{E}_\mu\left[f|s\right]$, the calculations are identical to those of Lemma 5.10. ■

5.3.3 Bound via Bernoulli-Laplace model, again

Lemma 5.13. *There exists a constant $k > 0$ not depending on n such that the first part of equation (5.40) is bounded by*

$$\mathbb{E}\left[\mathbb{E}_\mu\left[f|s\right] \cdot \mathrm{Ent}_\mu\left(\frac{f}{\mathbb{E}_\mu\left[f|s\right]}\middle| s\right)\right] \leq k\log(n) \cdot \mathcal{D}(f,f). \tag{5.50}$$

Proof: Also for this lemma we would like to apply the same method used for the bound of the spectral gap. Unfortunately some 'problems' arise in this case. In fact, in paper [15], T. Lee and H. Yau provided the following bounds for the log-Sobolev constant $\alpha^{BL}_{n,r}$ of the Bernoulli-Laplace model with n sites and r particles:

$$\varepsilon \log\frac{n^2}{n-r} \leq \frac{1}{\alpha^{BL}_{n,r}} \leq \frac{2}{\log 2}\log\frac{n^2}{n-r}, \tag{5.51}$$

where ε is a strictly positive constant independent of n and r. Hence, for very big or very small values of r (equivalently, for configurations with a very big or a very small number of "+"-spins) we cannot bound the log-Sobolev with a constant. Anyway, let's follow the path of the other proof, bounding the Bernoulli-Laplace log-Sobolev constant with its worst-case for all the values of the magnetization. Call

$$g := \frac{f}{\mathbb{E}_\mu [f|s]};$$

there exists a constant $c > 0$ not depending on n and $r = n_+$, such that

$$\begin{aligned}\mathrm{Ent}_\mu (g|s) = \mathrm{Ent}^{BL}_{n,n_+} (g) &\le \frac{1}{\alpha^{BL}_{n,n_+}} \mathcal{D}^{BL}_{n,n_+} (\sqrt{g}) \le (c \log n) \mathcal{D}^{BL}_{n,n_+} (\sqrt{g}) \\ &\le (c \log n) \frac{1}{2n} \sum_{i,j} \mathbb{E}_\pi \left[\left(\frac{\sqrt{f(\sigma^{ij})}}{\sqrt{\mathbb{E}_\mu [f|s]}} - \frac{\sqrt{f(\sigma)}}{\sqrt{\mathbb{E}_\mu [f|s]}} \right)^2 \middle| s \right]\end{aligned}$$

where $\pi(\sigma) = 1/\binom{n}{n_+}$ for all σ with $S(\sigma) = s$. Therefore

$$\mathbb{E}_\mu [f|s] \cdot \mathrm{Ent}_\mu \left(\frac{f}{\mathbb{E}_\mu [f|s]} \middle| s \right) \le (c \log n) \frac{1}{2n} \sum_{i,j} \mathbb{E}_\pi \left[\left(f(\sigma^{ij})^{\frac{1}{2}} - f(\sigma)^{\frac{1}{2}} \right)^2 \middle| s \right].$$

After that, we can take the expectation on both sides and then proceed exactly as in the proof of Lemma 5.9. We finally get

$$\mathbb{E} \left[\mathbb{E}_\mu [f|s] \cdot \mathrm{Ent}_\mu \left(\frac{f}{\mathbb{E}_\mu [f|s]} \middle| s \right) \right] \le k \log(n) \cdot \mathcal{D}(f, f)$$

as requested. ■

5.3.4 Conclusions

Plugging Lemma 5.12 and Lemma 5.13 in equation (5.40) we have a bound for the entropy of any function $f > 0$ such that $\mu(f) = 1$. The theorem follows at once.

Theorem 5.14. *For $\beta < 1$, there exists a constant $c > 0$ such that the log-Sobolev constant of the Glauber dynamics for the Ising model on the n-complete graph verifies*

$$\alpha \ge c \cdot \frac{1}{\log n}. \tag{5.52}$$

Anyway, this result doesn't give a satisfactory answer for the actual order of the log-Sobolev constant. Is $(\log n)^{-1}$ the correct behaviour? We couldn't find a test-function that confirms this hypothesis.

Most likely the result of Lemma 5.13 is not sharp. In its proof we took the worst case of the inverse of the log-Sobolev constant for the Bernoulli-Laplace model even when n_+ or n_- were not comparable with n, that is the most of the cases. Possibly we lost precision in that passage. Since an intermediate order between $O((\log n)^{-1})$ and $O(1)$ is hardly believable, it seems reasonable that α is just a constant not depending on n, as the spectral gap was. But this, of course, has to be verified.

Bibliography

[1] Aldous D. and Diaconis P. (1986) *Shuffling cards and stopping times*, Amer. Math. Monthly **93**, 333-348.

[2] Aldous D. and Fill J. (1999) *Reversible Markov chains and random walks on graphs*, in progress. Manuscript available at http://www.stat.berkeley.edu/∼aldous/RWG/book.html.

[3] Bodineau T. and Martinelli F. (2002) *Some new results on the kinetic Ising model in a pure phase*, arXiv:math-ph/0202013v1.

[4] Casella G. and Robert C. (2008) *A History of Markov Chain Monte Carlo*, http://www.stat.ufl.edu/∼casella/Papers/MCMCHistory.pdf.

[5] Chen G.-Y. and Saloff-Coste L. (2008) *The cutoff phenomenon for ergodic Markov processes*, Electronic Journal of Probability, Vol. 13, paper no. 3, 26-78, http://www.math.washington.edu/∼ejpecp.

[6] Diaconis P. (2008) *The Markov Chain Monte Carlo Revolution*, Bull. Amer. Math. Soc. Nov. 2008, http://www-stat.stanford.edu/∼cgates/PERSI/papers/MCMCRev.pdf.

[7] Diaconis P. and Saloff-Coste L. (2006) *Separation cut-offs for birth and death chains*, Ann. Appl. Probab. **16** , no.4, 2098-2122.

[8] Diaconis P. and Shahshahani M. (1987) *Time to reach stationarity in the Bernoulli-Laplace diffusion model*, SIAM J. Math. Anal. **18**, 208-218.

[9] Ding J., Lubetzky E. and Peres Y. (2008) *The mixing time evolution of Glauber dynamics for the mean-field Ising model*, arXiv:0806.1906v2.

[10] Ding J., Lubetzky E. and Peres Y. (2008) *Total variation cutoff in birth and death chains*, arXiv:0801.2625v4.

[11] Fill J. A. (2009) *The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof*, Journal of Theoretical Probability, to appear.

[12] Gao F. and Quastel J. (2003) *Exponential decay of entropy in the random transposition and Bernoulli-Laplace models*, Ann. Appl. Probab. **13**, no.4, 1591-1600.

[13] Griffiths R. B., Weng C.-Y. and Langer J. S. (1966) *Relaxation Times for Metastable States in the Mean-Field Model of a Ferromagnet*, Phys. Rev. **149**, 301-305.

[14] Karlin S. and McGregor J. (1959) *Coincidence properties of birth and death processes*, Pacific J. Math. **9**, 1109 -1140.

[15] Lee T. and Yau H. (1998) *Logarithmic Sobolev inequality for some models of random walks*, Annals of Probability, **26** , no.4, 1855-1873.

[16] Levin D. A., Malwina L. J. and Peres Y. (2007) *Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability*, arXiv:0712.0790v2.

[17] Levin D. A., Peres Y. and Wilmer E. L. (2009) *Markov Chains and Mixing Times*, American Mathematical Society.

[18] Mitzenmacher M., Upfal E. (2005) *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge.

[19] Norris J. R. (1998) *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge. Reprint of 1997 original.

[20] Reed M. and Simon B. (1978) *Analysis of Operators*, Methods of modern mathematical physics, Vol. 4, Academic Press, New York.

[21] Saloff-Coste L. (1997) *Lectures on finite Markov chains*, Lectures On Probability Theory and Statistics, Ecole d'Ete de Probabilites de Saint-Flour XXVI, 301-413.

[22] Shiryaev A. N. (1980) *Probability*, Graduate Texts in Mathematics, Springer.

[23] Williams D. (1991) *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University press, Cambridge.

Contents