



**HAL**  
open science

# On concentration inequalities for equilibrium states in lattice and symbolic dynamical systems

Jordan Moles

► **To cite this version:**

Jordan Moles. On concentration inequalities for equilibrium states in lattice and symbolic dynamical systems. Mathematical Physics [math-ph]. Institut Polytechnique de Paris; Universidad Autónoma de San Luis Potosí, 2020. English. NNT : 2020IPPAX102 . tel-03152537

**HAL Id: tel-03152537**

**<https://theses.hal.science/tel-03152537v1>**

Submitted on 25 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2020IPPAX102

Thèse de doctorat



# On concentration inequalities for equilibrium states in lattice and symbolic dynamical systems

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à Ecole polytechnique and Universidad autonoma de San Luis Potosi

École doctorale n°626 Ecole Doctorale de l'Institut Polytechnique de Paris (IPParis)  
Spécialité de doctorat : Physique

Thèse présentée et soutenue à Palaiseau, le 18 décembre 2020, par

**JORDAN MOLES**

Composition du Jury :

Frank Redig Professeur, Université de Delft, Pays-Bas	Président
Sandro Vaienti Professeur des universités, Aix-Marseille Université	Rapporteur
Aernout van Enter Professeur émérite, Université de Groningen, Pays-Bas	Rapporteur
Sandro Gallo, Professeur assistant, Université de Sao Carlos, Brésil	Examineur
Jean-René Chazottes Directeur de recherche, Ecole Polytechnique	Co-directeur de thèse
Edgardo Ugalde Professeur, Université autonome de San Luis Potosi, Mexique	Co-directeur de thèse



# Acknowledgements

First of all, I would like to express my deepest gratitude to my research supervisors Jean-René Chazottes from Ecole Polytechnique and Edgardo Ugalde from the Universidad Autonoma de San Luis Potosi who co-directed this thesis with patience and willingness. It was a great privilege to do research under their expertise and guidance.

I would like to pay my special regards to the president of the jury and professor Frank Redig for the enlightening collaboration and his great advice about the two dimensional Ising model.

I am indebted to professor Aernout van Enter for invaluable discussions which helped to refine this work and for accepting to be rapporteur.

I thank professors Sandro Vaienti and Sandro Gallo for accepting respectively to be rapporteur and member of the jury and for their helpful comments and suggestions to improve my thesis.

Another mention goes to my thesis committee Gelasio Salazar and Jesus Dorantes which encouraged my thesis project.

My acknowledgements also goes to the people at the secretary of the Instituto de Física and CPHT for helping me to solve administrative and migration issues.

During this four years, I enjoyed to learn in different and convivial working environments both in France and Mexico. I wish to thank specially all the people from “el grupo de dinámica potosina” and the Tartaglia seminar at Instituto de Física for enriching conversations (mathematical or not).

My thanks would not be complete if I did not mention my family and my relatives who encouraged me and supported me throughout these years.

**Titre :** Inégalités de concentration pour des états d'équilibre sur réseau et des systèmes dynamiques symboliques

**Mots clés :** Inégalités de concentration, formalisme thermodynamique, mesures de Gibbs, systèmes sur réseau, transition de phase, modèle d'Ising

**Résumé :** Cette thèse traite des propriétés de concentration d'états d'équilibre pour des systèmes sur réseau.

Dans le premier chapitre, on établit une relation entre transitions de phase sur un espace général de configurations et la perte de la concentration Gaussienne. Plus précisément, nous montrons que si un état d'équilibre associé à un potentiel invariant par décalage et absolument sommable satisfait la concentration Gaussienne alors il est à fortiori mélangeant et unique. On prouve ce théorème de deux manières différentes. L'une utilise les grandes déviations et l'autre est une conséquence d'un théorème plus abstrait qui dit que si une mesure de probabilité ergodique satisfait GCB alors elle a la propriété d'entropie relative inférieure positive.

Ensuite, dans le but de clarifier le lien entre ces concepts, nous étudions numériquement un modèle physique particulier autorisant une transition de phase. Comme candidat naturel, nous choisissons de simuler le modèle d'Ising ferromagnétique avec premiers voisins sans champ magnétique extérieur en dimension deux. Nous évaluons les constantes de la concentration grâce à la simulation d'observables classiques à toute température. Grâce au comportement de ces paramètres, nous mettons spécialement en lumière le comportement de concentration Gaussienne à toute température au dessus de la température critique. Nous observons notamment la perte de la concentration Gaussienne et nous analysons la divergence en loi de puissance de la constante de concentration Gaussienne à la température critique. Dans le but de compléter l'étude, nous quantifions les fluctuations d'observables d'intérêts dans le régime des basses températures dans lequel les états d'équilibre (ou phases) ne peuvent concentrer de la même manière. Nous renforçons aussi les résultats théoriques démontrés par J-R. Chazottes, P. Collet, F. Redig ou une borne de concentration stretched-exponential a été prouvé pour les phases en dessous de la température critique. Enfin, nous déterminons le comportement des constantes de concentration en fonction de la température.

Par la suite, motivés par les simulations du chapitre précédent, nous avons pour but de prouver que le

modèle d'Ising ferromagnétique avec ou sans champ magnétique extérieur en dimension deux satisfait une borne de concentration Gaussienne dans tout le régime d'unicité sauf à la température critique lorsque  $h = 0$ . La preuve se base sur plusieurs résultats connus que nous devons assembler. Pour ce modèle, nous utilisons le fait dans tout le régime d'unicité les mesures de Gibbs de volumes finis associés au potentiel satisfont la condition de weak mixing. Ensuite, on utilise un résultat général prouvé par F. Martinelli, E. Olivieri, and R. H. Schonmann qui dit que weak mixing est équivalent à strong mixing pour tous les carrés pour des systèmes de spins sur réseau en dimension deux. Cette condition implique que l'unique mesure de Gibbs en volume infini satisfait une inégalité logarithmique de Sobolev. Afin de terminer la preuve, nous prouvons que la propriété précédente implique que l'état d'équilibre satisfait une borne de concentration Gaussienne.

Nous dédions un chapitre à l'étude d'un système dynamique symbolique unidimensionnel sur un alphabet fini: les chaînes à liaisons complètes. Nous étudions en particulier les propriétés de concentration de l'unique état d'équilibre associé à un potentiel (ou probabilité de transition) non-Hölderien satisfaisant la condition de Walters. Pour des potentiels Hölderien par rapport à la distance classique ou pour des potentiels à variation exponentielle, nous savons qu'il existe un unique état d'équilibre exponentiellement mixing. De plus, Jean-René Chazottes et Sebastien Gouezel ont prouvé qu'il satisfait GCB. Nos résultats disent que GCB restent vrai pour une grande classe de potentiels  $\phi$  satisfaisant la condition de Walters qui inclut la condition de variation sous-exponentielle.

Enfin, nous traitons un autre système dynamique important que sont les automates cellulaires probabilistes de voisinage fini. Cette étude aborde les automates cellulaires probabilistes comme une perturbation d'automates cellulaires déterministes dans un régime de haut bruit dans lequel la mesure de Gibbs spatio-temporelle associée à la dynamique est l'unique mesure invariante par la dynamique et invariante par translation et a des propriétés de mélange exponentiel. Dans ce contexte, nous prouvons que cette mesure satisfait aussi une borne de concentration Gaussienne.

**Title :** On concentration inequalities for equilibrium states on lattice and symbolic dynamical systems

**Keywords :** Concentration inequalities, thermodynamic formalism, Gibbs measures, lattice systems, phase transition, Ising models

**Abstract :** This thesis deals with concentration properties of equilibrium states for lattice systems.

In the first chapter, we establish a relation between phase transitions on a general configuration space and the loss of the Gaussian concentration bound. More precisely, we show that if an equilibrium state associated to a shift-invariant and absolutely summable potential satisfies a Gaussian concentration bound then it is à fortiori mixing and unique. We prove this theorem in two different way. One uses large deviations and the other one is a consequence of a more abstract theorem which says that if an ergodic probability measure satisfies GCB then it has the lower positive relative entropy property.

Thereafter, in view of clarifying the link between these concepts, we study numerically a particular physical model which allows phase transition to occur. As a natural candidate, we chose to compute the nearest-neighbor ferromagnetic Ising model without external magnetic field in two dimensions. We evaluate concentration constants through classical estimates at all temperatures. Thank to the behavior of these parameters, at all temperatures above the critical one, we emphasize that the Gaussian concentration holds. We analyze the power-law divergence of the Gaussian concentration constant at the critical temperature and observe the loss of the Gaussian concentration. To complete the study, we quantify the fluctuations of observables of interest in the low-temperature regime in which the equilibrium states (or phases) cannot concentrate in the same way. We also reinforce the theoretical results proved by J-R. Chazottes, P. Collet, F. Redig where a stretched-exponential concentration bound was proven for the Gibbs measure below the critical temperature. To achieve this, we determined the behavior of the concentration parameters according to the temperature.

Later on and motivated by the simulations in the previous chapter, we aim to prove that for the 2D ferromagnetic Ising model with or without external magnetic

field the Gaussian concentration bound holds in the whole uniqueness regime except at the critical temperature when  $h = 0$ . The proof is based on several known results that we have to put together. For this model, we first use the fact that in the whole uniqueness regime the finite-volume Gibbs measures associated to the potential satisfies the weak mixing condition. Then, we use a general result proved by F. Martinelli, E. Olivieri, and R. H. Schonmann which says that weak mixing is equivalent to strong mixing for all squares for 2D lattice spin systems. This condition implies that the unique infinite-volume Gibbs measure satisfies a logarithmic Sobolev inequality. To complete the proof, we proved that the previous property implies that the equilibrium state satisfies a Gaussian concentration bound.

We dedicate a chapter to the study of an unidimensional symbolic dynamics on a finite alphabet: chains with complete connections. In particular, we study the concentration properties of a unique equilibrium state associated to a non-Hölderian potential (or transition probability) satisfying a Walters's condition. For Hölderian potentials with respect to the classical distance or for potentials with an exponential variation, we know that there exists an exponentially mixing unique equilibrium state. Moreover Jean-René Chazottes and Sebastien Gouezel proved that it satisfies GCB. Our results says that GCB remains true for a large subclass of potentials  $\phi$  satisfying Walters condition which includes the sub-exponential variation condition.

Finally, we deal with another important dynamical system which is probabilistic cellular automata with finite neighborhood. This study tackles the probabilistic cellular automata as a perturbation of the deterministic cellular automata in a high noise regime in which the Gibbs measure associated to the dynamics is the unique space-time invariant shift invariant measure and has exponential mixing properties. In this context, we prove that this measure satisfies also a Gaussian concentration bound.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Setting</b>	<b>5</b>
2.1	Configuration space . . . . .	5
2.2	Basics of ergodic theory . . . . .	5
2.3	Gibbs measures and equilibrium states . . . . .	8
<b>3</b>	<b>Concentration inequalities</b>	<b>13</b>
3.1	Basics of concentration inequalities . . . . .	13
3.2	Known results for Gibbs measures . . . . .	16
3.2.1	Gaussian concentration bound . . . . .	16
3.2.2	Moment concentration bounds . . . . .	20
3.2.3	Stretched-exponential concentration bound . . . . .	20
3.3	Known results for stochastic chains of unbounded memory . . . . .	22
3.4	On the relation between concentration and large deviations for equilibrium states . . . . .	23
<b>4</b>	<b>Gaussian concentration and uniqueness of equilibrium states in lattice systems</b>	<b>25</b>
4.1	Setting . . . . .	26
4.2	Proof of Theorem 4.0.1 . . . . .	29
4.2.1	An abstract result . . . . .	29
4.2.2	GCB implies the blowing-up property . . . . .	30
4.2.3	Blowing-up implies exponential rate for frequencies . . . . .	31
4.2.4	Exponential rate for frequencies implies positive relative entropy property . . . . .	33
4.3	Another proof of Theorem 4.0.1 . . . . .	34
4.4	Appendix . . . . .	35
4.4.1	Proof of Lemma 4.1.1 . . . . .	35
4.4.2	Proof of Lemma 4.2.1 . . . . .	35
4.4.3	A bound on relative entropy . . . . .	37
4.5	A final remark . . . . .	37
<b>5</b>	<b>Numerical study of concentration inequalities for the 2D Ising model</b>	<b>39</b>
5.1	Setting . . . . .	39
5.2	Computations and estimates . . . . .	40
5.2.1	Introduction . . . . .	40
5.2.2	Computation problems . . . . .	41
5.2.3	Classical Metropolis algorithm . . . . .	42
5.2.4	Finite-size effect . . . . .	43



5.2.5	Finite sampling time effect . . . . .	43
5.3	Estimation of concentration constants . . . . .	44
5.3.1	Method . . . . .	44
5.3.2	Settings and preliminary results . . . . .	45
5.3.3	Gaussian concentration constant estimation . . . . .	46
5.3.4	Determination of concentration constant behavior . . . . .	48
5.3.5	Stretched-exponential concentration constant estimation . . . . .	50
5.4	Conclusion . . . . .	57
5.5	Code . . . . .	57
<b>6</b>	<b>Gaussian concentration bound for the 2D Ising model</b>	<b>61</b>
6.1	Proof of Theorem 6.0.1 . . . . .	61
6.1.1	Weak mixing implies strong mixing for 2D lattice spin systems . . . . .	61
6.1.2	Strong mixing implies the logarithmic Sobolev inequality . . . . .	64
6.1.3	Logarithmic Sobolev inequality implies GCB . . . . .	65
6.1.4	Is Logarithmic Sobolev inequality equivalent to complete analyticity? . . . . .	66
6.2	Appendix: Proof of Lemma 6.1.1 . . . . .	66
<b>7</b>	<b>Gaussian concentration for potentials on <math>S^{\mathbb{N}}</math> with subexponential variations</b>	<b>69</b>
7.1	introduction . . . . .	69
7.2	Settings . . . . .	70
7.3	Speed of convergence of the transfer operator . . . . .	73
7.4	Gaussian concentration bound . . . . .	75
7.4.1	Related works . . . . .	77
7.5	Proof of Theorem 7.4.1 . . . . .	78
7.5.1	Some preparatory results . . . . .	78
7.5.2	Proof of Theorem 7.4.1 . . . . .	79
7.6	Applications . . . . .	82
7.6.1	Birkhoff sums . . . . .	82
7.6.2	Empirical frequency of blocks . . . . .	83
7.6.3	Hitting times and entropy . . . . .	85
7.6.4	Speed of convergence of the empirical measure . . . . .	85
7.6.5	Relative entropy, $\bar{d}$ -distance and speed of Markov approximation . . . . .	86
7.6.6	Shadowing of orbits . . . . .	89
7.6.7	Almost-sure central limit theorem . . . . .	90
7.7	Appendix . . . . .	92
7.7.1	Cones and projective metrics . . . . .	92
7.7.2	Proof V. Maume's theorem . . . . .	95
<b>8</b>	<b>Probabilistic cellular automata</b>	<b>103</b>
8.1	Introduction . . . . .	103
8.2	Setting . . . . .	103
8.3	High-noise regime . . . . .	107
8.4	Gaussian Concentration Bound . . . . .	110

<b>9 Perspectives</b>	<b>117</b>
9.1 Concentration inequalities for marginal distributions of the joint distribution of the PCA . . . . .	117
9.2 Coupled map lattices . . . . .	117
9.3 Countable Markov shifts . . . . .	119
9.4 Relationship between GCB and complete analyticity . . . . .	120
<b>Bibliography</b>	<b>121</b>



# Chapter 1

## Introduction

The phenomenon we are interested in, which goes under the name of “concentration inequalities”, is that if a function of many “weakly dependent” random variables does not depend too much on any of them, then it is concentrated around its expected value. Among the various areas of applications (see [10, 31, 53]), it plays an important role in probability theory, statistics and for our concern in statistical mechanics. Consider independent and identically distributed random variables  $\{\omega_x, x \in \Lambda_n\}$  taking values in  $\{-1, +1\}$ , where  $\Lambda_n = [-n, n]^d \cap \mathbb{Z}^d$  is the discrete cube of volume  $(2n + 1)^d$ . By the law of large numbers, the sum  $\frac{1}{(2n+1)^d} \sum_{x \in \Lambda_n} \omega_x$  converges to its expected value (in this case 0) and its fluctuations are obviously included in an interval of size  $\mathcal{O}(n^d)$ . In fact, by application of the well-know Hoeffding theorem, such an observable concentrates sharply around its mean in an interval of size  $\mathcal{O}(n^{d/2})$  with high probability:

$$\mathbb{P} \left( \left| \frac{1}{(2n+1)^{d/2}} \sum_{x \in \Lambda_n} \omega_x \right| > u \right) \leq 2 \exp \left( -\frac{u^2}{2} \right)$$

for all  $n \geq 1$  and for all  $u > 0$  (see [10]). When this bound holds, we say that the measure  $\mathbb{P}$  satisfies a Gaussian concentration inequality.

Besides, this phenomenon doesn't hold only for linear functions of  $\omega_x$ 's (like the sum above) but for a large class of non-linear functions  $F$  of the  $\omega_x$ 's. Another interesting feature of concentration inequalities is that, unlike central limit theorems or large deviation inequalities, they are non-asymptotic. By non-asymptotic we mean that we allow the number of parameters to be large but finite.

In the general case of dependent random variables, the situation is naturally more complex but one may expect to have a Gaussian concentration bound as above for weakly dependent random variables  $\omega_x$ 's (see [64, 63, 62, 74] for the Markovian case). In this thesis, we are interested in Gibbs measures on a configuration space of the form  $\Omega = S^{\mathbb{L}}$  where  $S$  is a finite set and  $\mathbb{L} = \mathbb{N}$  or  $\mathbb{Z}^d$  with  $d \geq 1$ . For  $\mathbb{L} = \mathbb{N}$ , R. Fernández and G. Mailard studied the link between one-dimensional Gibbs measures associated to a potential and the generalization of Markov chains known as “chains with complete connections” (see [36]). These chains are defined by conditional probabilities of observing the next symbol which may depend on the whole “past”. In symbolic dynamics, such discrete-time stochastic processes are known as “g-measures”. Although there exist equivalence statements under some regularity conditions, in general g-measures and Gibbs measures are

not equivalent. Actually, in [34], the authors constructed a non-Gibbsian g-measure and in [7], R. Bissacot E. Endo A. van Enter and A. Le Ny identified a Gibbs measure which is not a g-measure. In this case, we adopt a dynamical system approach and we exhibit conditions for which the associated unique Gibbs measures satisfies a Gaussian concentration inequality.

We will first consider Gibbs measures on  $d$ -dimensional lattices. In the previous example, the product measure can be identified as a Gibbs measure at infinite temperature on  $\{-1, +1\}^{\mathbb{Z}^d}$  where  $K$  is the magnetization inside the cube  $\Lambda_n$  namely  $\sum_{x \in \Lambda_n} \omega_x / |\Lambda_n|$ . This Gaussian concentration bound was first proved in [50] for various types of local function  $K$  and for potentials satisfying Dobrushin's uniqueness condition, with a constant explicitly related to Dobrushin's interdependence matrix. This covers, for instance, finite-range potentials at sufficiently high temperature. Not surprisingly, one cannot expect that a Gaussian concentration holds for the (ferromagnetic) Ising model at temperatures below the critical one, because of the surface-order large deviations of the magnetization (see [14] for more details). In [14], the authors proved that a "stretched-exponential" concentration bound holds for the "+" phase and the "-" phase of this model at sufficiently low temperature. In fact, we will prove that if there exist several equilibrium states for a potential, none of them can satisfy a Gaussian concentration bound [19].

In the ferromagnetic Ising model in two dimensions, we could quantify numerically the speed with which the Gaussian concentration constant diverges, when one approaches the critical point from above; this turns out to be a power law in the temperature. A similar behavior is found for the stretched-exponential concentration constant when one approaches the critical point from below. We simulated various type of observables  $K$  of the  $\omega_x$ 's at different temperatures and estimated the concentration constant associated.

This numerical study paved our way for proving that the Gaussian concentration bound holds for all temperatures above the critical one. Let us also mention that the proof relies on different mixing properties for the Gibbs state and is only valid for finite-range discrete spin systems on the two dimensional lattice.

We will also study the concentration properties of the one-sided shift on  $S^{\mathbb{N}}$  associated to a potential satisfying Walters condition with sub-exponential continuity rates. It is well-known that when the potential is Hölder with respect to the usual metric on the configuration space there exists a unique equilibrium state ([83, 11, 70]) and it satisfies a Gaussian concentration bound (see [17]). In this context, the Ruelle-Perron-Frobenius operator (or transfer operator) associated to the dynamics is quasi-compact and implies an exponential decay of correlations. But when the quasi-compactness doesn't hold anymore, we obtain a sub-exponential decay of correlation which implies the existence of a unique equilibrium state (see [66]). We will prove under a summability condition on the coefficient which controls the decay of correlation that a Gaussian concentration bound holds for the unique equilibrium state (see [20]).

Finally, we deal with another important dynamical system which is probabilistic cellular automata with finite neighborhood (see [58]). This study tackles the probabilistic cellular automata as a perturbation of the deterministic cellular automata in a high-noise regime ([52]) in which the Gibbs measure associated to the dynamics is the unique space-time invariant measure and

has exponential mixing properties. In this context, we prove that the spatio-temporal measure satisfies a Gaussian concentration bound.



# Chapter 2

## Setting

### 2.1 Configuration space

In this thesis, we are interested in models made of a configuration space of the form  $\Omega = S^{\mathbb{L}}$  where  $\mathbb{L} = \mathbb{Z}^d$  with  $d \geq 1$  or  $\mathbb{L} = \mathbb{N}$ , and  $S$  is a non-empty finite set. We endow  $\Omega$  with the product topology that is generated by cylinder sets. We denote by  $\mathfrak{B}$  the Borel  $\sigma$ -algebra which coincides with the  $\sigma$ -algebra generated by cylinder sets.

An element  $x$  of  $\mathbb{Z}^d$  (hereby called a site) can be written as a vector  $(x_1, \dots, x_d)$  in the canonical base of the lattice  $\mathbb{Z}^d$ . Let  $\|x\|_{\infty} = \max_{1 \leq i \leq d} |x_i|$  and  $\|x\|_1 = |x_1| + \dots + |x_d|$ . If  $\Lambda$  is a finite subset of  $\mathbb{Z}^d$ , denote by  $|\Lambda|$  its cardinality and  $\text{diam}(\Lambda) = \max\{\|x\|_{\infty}; x \in \Lambda\}$  its diameter.

We consider the following distance on  $\Omega$ : For  $\omega, \omega' \in \Omega$ , let

$$d_{\theta}(\omega, \omega') = \theta^k \quad \text{where } k = \min \{\|x\|_{\infty} : \omega_x \neq \omega'_x\}. \quad (2.1.1)$$

where  $\theta \in (0, 1)$  is some fixed number. This distance induces the product topology and it is well-known that  $\Omega$  is a compact metric space.

If  $\Lambda$  is a finite subset of  $\mathbb{Z}^d$ , we will write  $\Lambda \Subset \mathbb{Z}^d$ . For  $\Lambda \subset \mathbb{Z}^d$ , we denote by  $\Omega_{\Lambda}$  the projection of  $\Omega$  onto  $S^{\Lambda}$ . Accordingly, an element of  $\Omega_{\Lambda}$  is denoted by  $\omega_{\Lambda}$  and is viewed as a configuration  $\omega \in \Omega$  restricted to  $\Lambda$ . For  $\sigma, \eta \in \Omega$ , we denote by  $\sigma_{\Lambda} \eta_{\Lambda^c}$  the configuration which agrees with  $\sigma$  on  $\Lambda$  and with  $\eta$  on  $\Lambda^c$ . We denote by  $\mathfrak{B}_{\Lambda}$  the  $\sigma$ -algebra generated by the coordinate maps  $\omega \mapsto \omega_x$  when we restrict  $x$  to  $\Lambda$ . We need to define centered ‘‘cubes’’: for every  $n \in \mathbb{Z}_+$ , let

$$\Lambda_n = \{x \in \mathbb{Z}^d : -n \leq x_i \leq n, i = 1, 2, \dots, d\}.$$

Given  $\Lambda \subset \mathbb{Z}^d$ , an element  $p_{\Lambda}$  of  $S^{\Lambda}$  is called a pattern with shape  $\Lambda$ , or simply a pattern. We will write  $p_n$  instead of  $p_{\Lambda_n}$ . We will also consider elements of  $S^{\Lambda}$  as configurations restricted to  $\Lambda$ . We will simply write  $\omega$  instead of  $\omega_{\Lambda}$  since we will always precise to which set  $\omega$  belongs. A pattern  $p_n \in S^{\Lambda_n}$  determines a cylinder set  $[p_n] = \{\omega \in \Omega : \omega_{\Lambda_n} = p_n\}$ . More generally, given  $\Lambda \Subset \mathbb{Z}^d$  and  $C \subset S^{\Lambda}$ , let  $[C] = \{\omega \in \Omega : \pi_{\Lambda}(\omega) \in C\}$  where  $\pi_{\Lambda}$  is the projection from  $\Omega$  onto  $S^{\Lambda}$ .

### 2.2 Basics of ergodic theory

We recall basic results from [44]. We denote by  $\mathcal{M}(\Omega)$  the set of probability measures. Let us notice that since  $\Omega$  is compact, then  $\mathcal{M}(\Omega)$  is also compact



in the weak topology. Recall that in this setting,  $(\mu_n)_{n \in \mathbb{N}}$  weakly converges to  $\mu$  if for any  $\Lambda \in \mathbb{Z}^d$  and any  $C \subset S^\Lambda$  one has  $\mu_n([C]) \rightarrow \mu([C])$ . We recall that we define the shift action  $(T_x, x \in \mathbb{L})$  as: for each  $x \in \mathbb{L}$ ,  $T_x : \Omega \rightarrow \Omega$  and  $(T_x \omega)_y = \omega_{y-x}$ , for all  $y \in \mathbb{L}$ . This is a continuous map. In fact, if  $\mathbb{L} = \mathbb{N}$ ,  $(T_x, x \in \mathbb{L})$  corresponds to the non-invertible shift map whereas, if  $\mathbb{L} = \mathbb{Z}^d$ , it corresponds to the invertible one. For the sake of definiteness, we take  $\mathbb{L} = \mathbb{Z}^d$  in the rest of this chapter. Since we will only deal with probability measures, by “measure” we will always mean “probability measure”.

**Definition 2.2.1.**

A measure  $\mu \in \mathcal{M}(\Omega)$  is shift-invariant or translation-invariant if  $\mu \circ T_x^{-1} = \mu$  for all  $x \in \mathbb{Z}^d$ , i.e., if  $\mu(T_x^{-1}A) = \mu(A)$  for all  $A \in \mathfrak{B}$ . Equivalently

$$\forall x \in \mathbb{Z}^d, \forall f \in \mathcal{C}(\Omega) \quad \int f \circ T_x d\mu = \int f d\mu$$

where  $\mathcal{C}(\Omega)$  is the set of real-valued continuous functions on  $\Omega$ . We denote by  $\mathcal{M}_T(\Omega)$  the set of shift-invariant Borel probability measures on  $\Omega$ .

The existence of such measures is ensured by the Krylov-Bogolubov theorem [44] which states that if  $T : \Omega \rightarrow \Omega$  is continuous, then  $\mathcal{M}_T(\Omega) \neq \emptyset$ .

We introduce some definitions for the study of measure-preserving transformations and their basic properties.

**Definition 2.2.2.**

A measure  $\mu \in \mathcal{M}(\Omega)$  is ergodic if  $\mu(A) = 0$  or  $\mu(A) = 1$  for all  $A \in \mathfrak{B}$  such that  $T_x^{-1}A = A$  for all  $x \in \mathbb{Z}^d$ .

Ergodic measures are extreme points of  $\mathcal{M}_T(\Omega)$  (see [27] Proposition 5.6 p. 24); we denote them by  $\text{ex } \mathcal{M}_T(\Omega)$ . The fundamental result in ergodic theory is the following theorem.

**Theorem 2.2.1 (Ergodic theorem, [44]).**

Let  $\mu \in \mathcal{M}(\Omega)$ . For each  $f \in L^1_\mu$  the limit

$$\bar{f}(\omega) := \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} f(T_x \omega) \quad (2.2.1)$$

exists  $\mu$ -a.e. and in  $L^1_\mu$ . The function  $\bar{f}$  is  $T$ -invariant and for each  $T$ -invariant set  $A \subset \Omega$

$$\int_A \bar{f} d\mu = \int_A f d\mu. \quad (2.2.2)$$

Moreover, if  $\mu$  is ergodic, then

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} f(T_x \omega) = \int f d\mu, \quad \mu\text{-a.e.} \quad (2.2.3)$$

The set of full  $\mu$ -measure such that (2.2.3) holds depends on  $f$ . However, since  $\Omega$  is compact, the space  $\mathcal{C}(\Omega)$  is separable. Hence, when  $\mu$  is ergodic, there exists a measurable subset  $G_\mu \subset \Omega$  with  $\mu(G_\mu) = 1$  and such that (2.2.3) holds for all  $f \in \mathcal{C}(\Omega)$  and for all  $\omega \in G_\mu$ . (See [27, Proposition

5.9 p. 25] for a proof.) This fact can be reformulated more compactly by using empirical measures which are defined as follows: Given  $\omega \in \Omega$  and any  $n \in \mathbb{N}$ , let

$$\mathcal{E}_n(\omega) := \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} \delta_{T_x \omega}. \quad (2.2.4)$$

Then, for  $\mu$ -almost every  $\omega$ , one has

$$\mathcal{E}_n(\omega) \xrightarrow{n \rightarrow \infty} \mu \quad (2.2.5)$$

where the convergence is in the weak topology on the space of probability measures  $\mathcal{M}(\Omega)$ .

One important characterization of the ergodic theorem is the following.

**Proposition 2.2.1 ([44]).**

For  $\mu \in \mathcal{M}(\Omega)$  the following assertions are equivalent:

1.  $\mu$  is ergodic.
2. For all  $f, h \in L^2_\mu$

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} \int (f \circ T_x) \cdot h \, d\mu = \int f \, d\mu \cdot \int h \, d\mu. \quad (2.2.6)$$

**Definition 2.2.3 (Mixing).**

A measure  $\mu \in \mathcal{M}(\Omega)$  is mixing if  $\forall A, B \in \mathfrak{B}$  the following holds

$$\forall \varepsilon > 0, \exists n > 0, \forall x \in \mathbb{Z}^d \setminus \Lambda_n, \quad |\mu(T_x^{-1}A \cap B) - \mu(A)\mu(B)| < \varepsilon. \quad (2.2.7)$$

Equivalently, one has

$$\lim_{\|x\|_\infty \rightarrow \infty} \mu(T_x^{-1}A \cap B) = \mu(A)\mu(B). \quad (2.2.8)$$

**Remark 2.1.** An immediate consequence is that mixing implies ergodicity.

**Proposition 2.2.2 ([44]).**

For  $\mu \in \mathcal{M}(\Omega)$  the following assumptions are equivalent:

1.  $\mu$  is mixing.
2. For all  $f, h \in L^2_\mu$

$$\lim_{\|x\|_\infty \rightarrow \infty} \int (f \circ T_x) \cdot h \, d\mu = \int f \, d\mu \cdot \int h \, d\mu. \quad (2.2.9)$$

3. For all  $f \in L^2_\mu$

$$\lim_{\|x\|_\infty \rightarrow \infty} \int (f \circ T_x) \cdot f \, d\mu = \left( \int f \, d\mu \right)^2. \quad (2.2.10)$$

## 2.3 Gibbs measures and equilibrium states

We refer to [39] or [37] for details. We consider shift-invariant absolutely summable potentials. More precisely, a potential is a family of functions  $(\Phi(\Lambda, \cdot))_{\Lambda \in \mathbb{Z}^d}$  such that, for each (nonempty)  $\Lambda \in \mathbb{Z}^d$ , the function  $\Phi_\Lambda : \Omega \rightarrow \mathbb{R}$  is  $\mathfrak{B}_\Lambda$ -measurable which simply means that the value of  $\Phi_\Lambda(\omega)$  is determined by  $\omega_\Lambda$ , the restriction of  $\omega$  to  $\Lambda$ . By shift-invariance we mean that  $\Phi(\Lambda + x, T_x \omega) = \Phi(\Lambda, \omega)$  for all  $\Lambda \in \mathbb{Z}^d$ ,  $\omega \in \Omega$  and  $x \in \mathbb{Z}^d$  (where  $\Lambda + x = \{y + x : y \in \Lambda\}$ ). Uniform summability is the property that

$$\sum_{\substack{\Lambda \in \mathbb{Z}^d \\ 0 \in \Lambda}} \|\Phi(\Lambda, \cdot)\|_\infty < \infty. \quad (2.3.1)$$

We shall denote by  $\mathcal{B}_T$  the space of shift-invariant uniformly summable potentials. An important subspace of  $\mathcal{B}_T$  is the space of finite-range potentials. Finite-range means that there exists  $R > 0$  such that  $\Phi(\Lambda, \omega) = 0$  if  $\text{diam}(\Lambda) > R$ . The smallest such  $R$  is called the range of the potential. The space of potentials of range  $R$  is denoted by  $\mathcal{B}_R$ , and the set of all finite-range potentials is dense in  $\mathcal{B}_T$ .

Let  $\Phi \in \mathcal{B}_T$  and  $\Lambda \in \mathbb{Z}^d$ , the associated Hamiltonian in the finite volume  $\Lambda$  with boundary condition  $\eta \in \Omega$  is given by

$$\mathcal{H}_\Lambda(\omega|\eta) = \sum_{\substack{\Lambda' \in \mathbb{Z}^d \\ \Lambda' \cap \Lambda \neq \emptyset}} \Phi(\Lambda', \omega_\Lambda \eta_{\mathbb{Z}^d \setminus \Lambda}). \quad (2.3.2)$$

The corresponding specification is defined as the family of probability kernels  $\gamma^\Phi = \{\gamma_\Lambda^\Phi\}_{\Lambda \in \mathbb{Z}^d}$  such that

$$\gamma_\Lambda^\Phi(\omega|\eta) = \frac{\exp(-\mathcal{H}_\Lambda(\omega|\eta))}{Z_\Lambda(\eta)} \quad (2.3.3)$$

where  $Z_\Lambda(\eta)$  is the normalizing factor commonly called partition function in  $\Lambda$ . We say that  $\mu$  is a Gibbs measure for the potential  $\Phi$  if  $\gamma_\Lambda^\Phi(\omega|\cdot)$  is a version of the conditional probability  $\mu(\omega_\Lambda | \mathfrak{B}_{\Lambda^c})$ . Equivalently, this means that for all  $A \in \mathfrak{B}$ ,  $\Lambda \in \mathbb{Z}^d$ , one has the so-called ‘‘DLR equations’’ (in honor of Dobrushin, Lanford and Ruelle)

$$\mu(A) = \int d\mu(\eta) \sum_{\omega' \in S^\Lambda} \gamma_\Lambda^\Phi(\omega'|\eta) \mathbb{1}_A(\omega'_\Lambda \eta_{\Lambda^c}). \quad (2.3.4)$$

We denote by  $\mathcal{G}(\Phi)$  the set of Gibbs measures for a given absolutely summable potential  $\Phi$ . This set is never empty, but it may be not reduced to a singleton. Let  $\mathcal{G}_T(\Phi) = \mathcal{G}(\Phi) \cap \mathcal{M}_T(\Omega)$ , that is, the set of shift-invariant probability measures for  $\Phi$ . This set is a Choquet simplex and may contain several (extremal) elements. We denote by  $\text{ex } \mathcal{G}_T(\Phi)$  the set of extreme points which coincides with the set of ergodic Gibbs measures for  $\Phi$ , that is,  $\text{ex } \mathcal{G}_T(\Phi) = \mathcal{G}_T(\Phi) \cap \text{ex } \mathcal{M}_T(\Omega)$ . Of course, when  $\mathcal{G}(\Phi)$  is a singleton, then the unique Gibbs measure is shift-invariant and ergodic.

We now define relative entropy. Let  $\mu, \nu \in \mathcal{M}_T(\Omega)$ . For each  $n \in \mathbb{N}$ , let

$$H_n(\nu|\mu) := \sum_{p_n \in S^{\Lambda_n}} \nu([p_n]) \log \frac{\nu([p_n])}{\mu([p_n])}.$$

We denote by  $\log$  the natural logarithm.

**Definition 2.3.1.**

Let  $\mu, \nu \in \mathcal{M}_T(\Omega)$ . The lower and upper relative entropies of  $\nu$  with respect to  $\mu$  are defined by

$$h_*(\nu|\mu) = \liminf_{n \rightarrow \infty} \frac{H_n(\nu|\mu)}{(2n+1)^d}, \quad h^*(\nu|\mu) = \limsup_{n \rightarrow \infty} \frac{H_n(\nu|\mu)}{(2n+1)^d}. \quad (2.3.5)$$

When the limit exists, we put  $h(\nu|\mu) = h_*(\nu|\mu) = h^*(\nu|\mu)$ .

It is well known that  $h_*(\nu|\mu)$  and  $h^*(\nu|\mu)$  are non-negative numbers. Note that  $H_n(\nu|\mu)$  can be  $+\infty$  if  $\mu([p_n]) = 0$  for some  $p_n$  but this will not happen when  $\mu$  is fully supported. Let  $\Phi \in \mathcal{B}_T$  and  $\mu \in \mathcal{G}(\Phi)$ . It is proved in [39] that for any  $\nu \in \mathcal{M}_T(\Omega)$  we have

$$h_*(\nu|\mu) = h^*(\nu|\mu) = h(\nu|\mu) = P(\Phi) + \sum_{0 \in \Lambda} |\Lambda|^{-1} \int \Phi(\Lambda, \cdot) d\nu - h(\nu) \quad (2.3.6)$$

where  $h(\nu)$  is the entropy of  $\nu$  and  $P(\Phi)$  is the pressure of  $\Phi$  defined as follows

$$h(\nu) := \lim_{n \rightarrow \infty} -\frac{1}{(2n+1)^d} \sum_{p_n \in S^{\Lambda_n}} \nu([p_n]) \log \nu([p_n])$$

$$P(\Phi) := \sup_{\nu \in \mathcal{M}_T(\Omega)} \left( h(\nu) + \sum_{0 \in \Lambda} |\Lambda|^{-1} \int \Phi(\Lambda, \cdot) d\nu \right).$$

Notice that,  $\nu \in \mathcal{M}_T(\Omega)$ ,  $h(\nu|\mu)$  is the same number for all  $\mu \in \mathcal{G}_T(\Phi)$ , so it is natural to define, for each  $\nu \in \mathcal{M}_T(\Omega)$ ,

$$h(\nu|\Phi) := h(\nu|\mu) \quad \text{where } \mu \text{ is any element in } \mathcal{G}_T(\Phi).$$

We recall the definition of equilibrium states.

**Definition 2.3.2** (Equilibrium states).

Let  $\Phi \in \mathcal{B}_T$ . A shift-invariant probability measure  $\nu$  such that  $h(\nu|\Phi) = 0$  is called an equilibrium state for  $\Phi$ .

We have the following fundamental result which is usually referred to as the variational principle for equilibrium states (see [39], Chapter 15, [72], Theorem 4.2).

**Theorem 2.3.1** (Variational principle).

Let  $\Phi \in \mathcal{B}_T$ . We have  $h(\nu|\Phi) = 0$  if and only if  $\nu \in \mathcal{G}_T(\Phi)$ . In particular,  $\mathcal{G}_T(\Phi)$  coincides with the set of equilibrium states for  $\Phi$ .

**Nearest-neighbor Ising model.** First consider the nearest-neighbor ferromagnetic Ising model ( $S = \{-1, +1\}$ ) with  $d \geq 2$ :

$$\Phi(\Lambda, \omega) = \begin{cases} -h\omega_x & \text{if } \Lambda = \{x\} \\ -J\omega_x\omega_y & \text{if } \Lambda = \{x, y\} \text{ and } \|x - y\|_1 = 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.3.7)$$

where  $J, h \in \mathbb{R}$  represents respectively the interaction strength and the external magnetic field. We will always assume that  $J > 0$  (the ferromagnetic case of the Ising model). Consider the potential  $\beta\Phi$  where  $\beta \in \mathbb{R}_+$  is the inverse temperature and the Gibbs measures  $\mu_{\beta\Phi}^+$  and  $\mu_{\beta\Phi}^-$  obtained as the infinite-volume limits of the corresponding specification with “+” and the “-” boundary conditions, respectively. There exists  $\beta_c = \beta_c(d) > 0$  such that, if  $\beta < \beta_c$  then it is well-known that  $\mu_{\beta\Phi} = \mu_{\beta\Phi}^+ = \mu_{\beta\Phi}^-$  is the unique ergodic Gibbs measure and if  $\beta > \beta_c$  then  $\mu_{\beta\Phi}^+$  and  $\mu_{\beta\Phi}^-$  are distinct ergodic Gibbs measures. For  $d = 2$ , for all  $\beta > \beta_c$ ,  $\mathcal{G}_T(\beta\Phi) = \{\lambda\mu_{\beta\Phi}^+ + (1-\lambda)\mu_{\beta\Phi}^- : \lambda \in [0, 1]\}$  (hence  $\text{ex } \mathcal{G}_T(\beta\Phi) = \{\mu_{\beta\Phi}^+, \mu_{\beta\Phi}^-\}$ ).

**Long-range Ising model.** Consider the long-range Ising model where  $S = \{-1, +1\}$ ,  $d = 1$ , and

$$\Phi(\Lambda, \omega) = \begin{cases} -\frac{\omega_x \omega_y}{|x-y|^\alpha} & \text{if } \Lambda = \{x, y\} \text{ such that } x \neq y \\ 0 & \text{otherwise} \end{cases}$$

with  $\alpha > 1$ . If  $1 < \alpha \leq 2$  then, there exists  $\beta_c > 0$  such that, for all  $\beta < \beta_c$ ,  $\mu_{\beta\Phi}^+ \neq \mu_{\beta\Phi}^-$  and  $\text{ex } \mathcal{G}(\beta\Phi) = \{\mu_{\beta\Phi}^+, \mu_{\beta\Phi}^-\}$ . We refer to [68] for the relevant references.

**Potts model.** Consider the nearest-neighbor Potts ferromagnet model for which  $S = \{1, \dots, N\}$  for sufficiently large  $N \in \mathbb{N}$ :

$$\Phi(\Lambda, \omega) = \begin{cases} J \mathbb{1}_{\omega_x = \omega_y} & \text{if } \Lambda = \{x, y\} \text{ and } \|x - y\|_1 = 1 \\ 0 & \text{otherwise} \end{cases}$$

where  $J > 0$  is the interaction strength. There exists  $0 < \beta_N < +\infty$  such that  $|\text{ex } \mathcal{G}(\beta_N\Phi)| = N + 1$ , one of these measures being symmetric under spin flip,  $|\text{ex } \mathcal{G}(\beta\Phi)| = N$  when  $\beta > \beta_N$  and when  $\beta < \beta_N$ ,  $|\text{ex } \mathcal{G}(\beta\Phi)| = 1$ .

**Dobrushin’s uniqueness regime.** Let  $\Phi \in \mathcal{B}_T$  and  $\gamma^\Phi$  be the corresponding specification. One says that  $\gamma^\Phi$  satisfies Dobrushin uniqueness condition if

$$\mathfrak{c}(\gamma^\Phi) := \sum_{x \in \mathbb{Z}^d} C_{0,x}(\gamma^\Phi) < 1 \quad (2.3.8)$$

where

$$C_{x,y}(\gamma^\Phi) := \sup \left\{ \|\gamma_{\{x\}}^\Phi(\cdot | \omega) - \gamma_{\{x\}}^\Phi(\cdot | \omega')\|_\infty : \omega, \omega' \text{ differ only at site } y \right\}.$$

Under this abstract condition valid in more general cases, there exists a unique equilibrium state for  $\Phi$  which we denote by  $\mu_\Phi$  (see [39], chapter 8 and [14]). An interesting sufficient condition for 2.3.8 to hold is that

$$\sum_{0 \in \Lambda} (|\Lambda| - 1) \delta(\Phi(\Lambda, \cdot)) < 2 \quad (2.3.9)$$

where

$$\delta(\Phi(\Lambda, \cdot)) = \sup_{\omega, \omega' \in \Omega} |\Phi(\Lambda, \omega) - \Phi(\Lambda, \omega')|.$$

Condition (2.3.9) is satisfied by any finite-range potential  $\beta\Phi$  provided that  $\beta$  is small enough. This defines a “high-temperature” regime. When  $\beta$  is

sufficiently small (high-temperature regime of the potential), we observe that  $\gamma^{\beta\Phi}$  satisfies Dobrushin uniqueness condition.

We now consider any potential such that  $\Phi(\{x\}, \omega) = -h\omega_x$  for all  $x \in \mathbb{Z}^d$  and  $h \in \mathbb{R}$  (external magnetic field). If  $h$  is sufficiently large, then 2.3.9 also holds. In fact the condition implying 2.3.9 reads

$$\exp\left(\frac{1}{2} \sum_{0 \in \Lambda: |\Lambda| > 1} \delta(\Phi(\Lambda, \cdot))\right) \sum_{0 \in \Lambda} (|\Lambda| - 1) \delta(\Phi(\Lambda, \cdot)) < e^{|h|}.$$

Finally, one can also check that this condition holds at low temperatures for potentials with unique ground state, *e.g.*, the nearest-neighbor Ising model with non zero external magnetic field and sufficiently large  $\beta$ , or for large enough  $|h|$  and for all  $\beta$ .



# Chapter 3

## Concentration inequalities

### 3.1 Basics of concentration inequalities

We start by reviewing elementary facts about concentration inequalities. Let  $(\mathcal{X}, \mathfrak{A}, \mathbb{P})$  be a probability space. We give several well-known inequalities to estimate how a real-valued random variable  $X$  deviates from its expected value  $\mathbb{E}(X)$ , *i.e.*, on how we can bound  $\mathbb{P}(X - \mathbb{E}(X) \geq u)$  and  $\mathbb{P}(X - \mathbb{E}(X) \leq -u)$  depending on the deviation  $u > 0$ .

An elementary but powerful device to quantify tail probabilities is Markov's inequality.

**Theorem 3.1.1** (Markov's inequality).

Let  $Z$  be a non-negative real-valued random variable and suppose that  $\mathbb{E}(Z) < +\infty$ . Then,

$$\forall u > 0, \mathbb{P}(Z \geq u) \leq \frac{\mathbb{E}(Z)}{u}. \quad (3.1.1)$$

A classical and easy improvement of this inequality can be achieved under stronger integrability conditions on  $Z$  as follows.

**Corollary 3.1.1.**

Let  $\phi$  be a non-negative and non-decreasing function defined on an interval  $I \subset \mathbb{R}$  and  $Z$  a real-valued random variable taking values in  $I$ . Then,

$$\forall u > 0, \mathbb{P}(Z \geq u) \leq \frac{\mathbb{E}(\phi(Z))}{\phi(u)}. \quad (3.1.2)$$

Chebyshev's inequality follows directly from this corollary by taking  $\phi(u) = u^2$  and  $Z = |X - \mathbb{E}(X)|$ :

$$\forall u > 0, \mathbb{P}(|X - \mathbb{E}(X)| \geq u) \leq \frac{\text{Var}(X)}{u^2}.$$

This result applies in various cases requiring weak properties of random variables and it is a sufficient condition to prove for instance the weak law of large numbers. For example, let's consider independent and identically distributed random variables  $(X_i)_{1 \leq i \leq n}$  and  $Z = X_1 + \dots + X_n - n\mathbb{E}(X_1)$  such



that  $\text{Var}(X_1) < +\infty$ . In this case, Chebyshev's inequality becomes

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_1)\right| \geq u\right) \leq \frac{\sigma^2}{nu^2}.$$

where  $\sigma^2 = \text{Var}(X_1)$ . If one can prove the existence of moments of higher order  $p > 2$ , then by setting  $\phi(u) = u^p$ , we would have a better estimate, namely

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq u) \leq \frac{\mathbb{E}(|Z - \mathbb{E}(Z)|^p)}{u^p}, \quad u > 0.$$

If the random variable  $Z$  is such that  $\mathbb{E}(|Z|^p) < +\infty$  for all  $p > 0$ , then one may choose the value of  $p$  which optimizes the upper bound. One can get an improvement by using  $\phi(u) = e^{\lambda u}$  where  $\lambda > 0$ . This is known as Chernoff's bounding method.

**Proposition 3.1.1** (Chernoff's inequality).

Let  $Z$  be a real-valued random variable such that  $\mathbb{E}(e^{\lambda Z}) < +\infty$  for all  $\lambda > 0$ . Then, for all  $u > 0$

$$\mathbb{P}(Z \geq u) \leq e^{-\lambda u} \mathbb{E}(e^{\lambda Z}). \quad (3.1.3)$$

Hoeffding inequality is an application of Chernoff's inequality for sums of independent *bounded* random variables.

**Theorem 3.1.2** (Hoeffding's inequality).

Let  $n \geq 1$  and  $(X_k)_{1 \leq k \leq n}$  be a sequence of independent real-valued random variables such that, for all sequences  $(a_k)_{1 \leq k \leq n}$ ,  $(b_k)_{1 \leq k \leq n}$  of real numbers with  $a_k < b_k$  for all  $k \leq n$ , we have

$$\forall k, \mathbb{P}(a_k \leq X_k \leq b_k) = 1. \quad (3.1.4)$$

Letting

$$S_n = \sum_{k=1}^n X_k \quad (3.1.5)$$

we have, for all  $u > 0$ ,

$$\mathbb{P}(S_n - \mathbb{E}(S_n) > u) \leq \exp\left(-\frac{2u^2}{\sum_{k=1}^n (b_k - a_k)^2}\right), \quad (3.1.6)$$

$$\mathbb{P}(S_n - \mathbb{E}(S_n) < -u) \leq \exp\left(-\frac{2u^2}{\sum_{k=1}^n (b_k - a_k)^2}\right), \quad (3.1.7)$$

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| > u) \leq 2 \exp\left(-\frac{2u^2}{\sum_{k=1}^n (b_k - a_k)^2}\right). \quad (3.1.8)$$

Subsequently, Hoeffding and Azuma proved that the following generalization holds for martingales with bounded differences. Given a filtration  $\mathcal{F} = (\mathcal{F}_0 = \{\Omega, \emptyset\} \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n)$ , let  $Z$  be a  $\mathcal{F}_n$ -measurable real-valued random variable. The main idea consists in writing  $Z$  as a telescopic sum of

reverse martingale differences, *i.e.*,

$$Z - \mathbb{E}[Z] = \sum_{i=1}^n D_i \quad (3.1.9)$$

where

$$D_i = \mathbb{E}[Z|\mathcal{F}_i] - \mathbb{E}[Z|\mathcal{F}_{i-1}]. \quad (3.1.10)$$

Let's define  $d_i := \sup_i D_i - \inf_i D_i$ . We have the following result.

**Theorem 3.1.3** (Azuma-Hoeffding inequality).

Let  $Z$  be a martingale with respect to a filtration  $\mathcal{F} = (\mathcal{F}_0 = \{\mathcal{X}, \emptyset\} \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n)$ . Then, for all  $\lambda \in \mathbb{R}$ ,

$$\mathbb{E}[\exp(\lambda(Z - \mathbb{E}(Z)))] \leq \exp\left(\frac{\lambda^2}{8} \sum_{i=1}^n d_i^2\right). \quad (3.1.11)$$

In particular, for all  $u > 0$ ,

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > u) \leq 2 \exp\left(-\frac{2u^2}{\sum_{k=1}^n d_k^2}\right). \quad (3.1.12)$$

**Remark 3.1.** One can always bound  $|d_i|$  by  $2\|D_i\|_\infty$ .

This Gaussian inequality has paved the way for the study of more general functions than the sum of bounded random variables. The first step was realized by McDiarmid who generalized Hoeffding's inequality 3.1.2 considering sufficiently "smooth" functions of independent random variables.

**Theorem 3.1.4** (McDiarmid's inequality).

Let  $n \geq 1$ ,  $(X_k)_{1 \leq k \leq n}$  be a sequence of independent random variables taking values in some space  $\mathcal{X}$  and  $K : \mathcal{X}^n \rightarrow \mathbb{R}$  be a function. Assume that there exists nonnegative constants  $\ell_i$  with  $1 \leq i \leq n$  such that for all  $x_1, \dots, x_n, x'_i \in \mathcal{X}$ , one has

$$|K(x_1, \dots, x_i, \dots, x_n) - K(x_1, \dots, x'_i, \dots, x_n)| \leq \ell_i.$$

Then, for all  $\lambda \in \mathbb{R}$ ,

$$\mathbb{E}[\exp(\lambda(K(X_1, \dots, X_n) - \mathbb{E}[K(X_1, \dots, X_n)]))] \leq \exp\left(\frac{\lambda^2}{8} \sum_{i=1}^n \ell_i^2\right).$$

In particular, for all  $u > 0$ ,

$$\mathbb{P}(|K(X_1, \dots, X_n) - \mathbb{E}[K(X_1, \dots, X_n)]| > u) \leq 2 \exp\left(-\frac{2u^2}{\sum_{k=1}^n \ell_k^2}\right).$$

**Remark 3.2.** The regularity condition on  $K$  known as the bounded difference property makes possible the application of McDiarmid's inequality to non-bounded random variables.

Inequality 3.1.4 is the prototype of what we are going to call Gaussian concentration bound.

Beyond the independent case, Markov chains were considered by K. Marton [62, 63, 64] and by P-M. Samson [74] where Gaussian concentration bounds hold. We also refer to the book of R. Douc, E. Moulines, P. Priouret and P. Soulier [30] for further details. Another setting where people proved Gaussian concentration bounds for partially Lipschitz functions are hyperbolic dynamical systems [17]. They showed that for sufficiently mixing invariant measures, GCB holds. In particular, for Young towers, they deduced Gaussian/polynomial concentration from the tails of the return time on the basis of the tower.

In this thesis, we will pay attention to the dependent aspect of bounded random variables distributed according to Gibbs measures. Indeed, we will discard the unbounded aspect of random variables where GCB may not hold even for sums. One of our motivations is to understand the loss of the Gaussian concentration bound when the random variables become highly dependent (for example in low temperature Ising model where we can't control uniformly  $\|D_i\|_\infty$  because of the correlations).

## 3.2 Known results for Gibbs measures

Recall that  $\Omega = S^{\mathbb{Z}^d}$  where  $S$  is a finite set. In this context, let  $K : \Omega \rightarrow \mathbb{R}$  be a continuous function and  $x \in \mathbb{Z}^d$ . The oscillation of  $K$  at  $x$  is defined by

$$\delta_x(K) = \sup \{|K(\omega) - K(\omega')| : \omega, \omega' \in \Omega \text{ differ only at site } x\}.$$

This is a quite natural object since, given  $\Lambda \Subset \mathbb{Z}^d$  and two configurations  $\omega, \eta \in \Omega$  such that  $\omega_{\Lambda^c} = \eta_{\Lambda^c}$ , one has

$$|K(\omega) - K(\eta)| \leq \sum_{x \in \Lambda} \delta_x(K) \mathbb{1}_{\{\omega_x \neq \eta_x\}}.$$

We shall say that  $K : \Omega \rightarrow \mathbb{R}$  is a *local function* if there exists  $\Lambda \Subset \mathbb{Z}^d$  (the dependence set of  $K$ ) such that for all  $\omega, \tilde{\omega}, \hat{\omega}$ ,  $K(\omega_\Lambda \tilde{\omega}_{\Lambda^c}) = K(\omega_\Lambda \hat{\omega}_{\Lambda^c})$ . (This means that  $K$  is  $\mathfrak{B}_\Lambda$ -measurable.) Equivalently, this means that  $\delta_x(K) = 0$  for all  $x \notin \Lambda$ . It is understood that  $\Lambda$  is the smallest such set. Local functions are continuous, hence bounded. (In fact, continuous functions are obtained as uniform limits of local functions.)

We write  $\underline{\delta}(K)$  for the infinite array  $(\delta_x(K), x \in \mathbb{Z}^d)$ , and let

$$\|\underline{\delta}(K)\|_2^2 := \|\delta(K)\|_{\ell^2(\mathbb{Z}^d)}^2 = \sum_{x \in \mathbb{Z}^d} \delta_x(K)^2.$$

It is convenient to define the following set of all local functions. Let

$$\mathcal{L} = \bigcup_{\Lambda \Subset \mathbb{Z}^d} \mathcal{L}_\Lambda \tag{3.2.1}$$

where  $\mathcal{L}_\Lambda$  is the set of local functions whose dependence set is  $\Lambda$ .

### 3.2.1 Gaussian concentration bound

We can now define what we mean by a Gaussian concentration bound.

**Definition 3.2.1** (Gaussian concentration bound).

Let  $\mu$  be a probability measure on  $(\Omega, \mathfrak{B})$ . We say that  $\mu$  satisfies a Gaussian concentration bound if there exists  $D = D(\mu) > 0$  such that, for all functions  $K \in \mathcal{L}$ , we have

$$\mathbb{E}_\mu[\exp(K - \mathbb{E}_\mu[K])] \leq \exp\left(D\|\underline{\delta}(K)\|_2^2\right). \quad (3.2.2)$$

For the sake of brevity, we say that  $\mu$  satisfies  $\text{GCB}(D)$ .

We denote by  $\mathbb{E}_\mu$  the expectation with respect to  $\mu$ . A key point in this definition is that  $D$  is independent of  $K$ , in particular it is independent of the size of the dependence set of  $K$ . Inequality (3.2.2) easily implies the following tail inequality that we will use several times.

**Proposition 3.2.1.**

If a probability measure  $\mu$  on  $(\Omega, \mathfrak{B})$  satisfies  $\text{GCB}(D)$  then, for all  $K \in \mathcal{L}$  and for all  $u > 0$ ,

$$\mu\{\omega \in \Omega : K(\omega) \geq \mathbb{E}_\mu[K] + u\} \leq \exp\left(-\frac{u^2}{4D\|\underline{\delta}(K)\|_2^2}\right). \quad (3.2.3)$$

*Proof.* If  $K \in \mathcal{L}$  then, for any  $\lambda > 0$ ,  $\lambda K$  obviously belongs to  $\mathcal{L}$ . Applying Markov's inequality and (3.2.2) we get

$$\begin{aligned} \mu\{\omega \in \Omega : K(\omega) - \mathbb{E}_\mu[K] \geq u\} &\leq \exp(-\lambda u) \mathbb{E}_\mu[\exp(\lambda(K - \mathbb{E}_\mu[K]))] \\ &\leq \exp\left(-\lambda u + D\|\underline{\delta}(K)\|_2^2 \lambda^2\right). \end{aligned}$$

Minimizing over  $\lambda$  yields (3.2.3). □

Thus, looking for a Gaussian concentration bound consists in finding the existence of a strictly positive constant  $D$  satisfying (3.2.2). In practice, getting the optimal constant is out of reach.

**Proposition 3.2.2.**

If a probability measure  $\mu$  on  $(\Omega, \mathfrak{B})$  satisfies  $\text{GCB}(D)$ , then it satisfies the variance inequality

$$\text{Var}_\mu(K) := \mathbb{E}_\mu[K^2] - (\mathbb{E}_\mu[K])^2 \leq 2D\|\underline{\delta}(K)\|_2^2 \quad (3.2.4)$$

for all functions  $K \in \mathcal{L}$ .

The proof goes roughly as follows, the details are left to the reader. Take  $\lambda > 0$  and apply (3.2.2) to  $\lambda K$ , subtract 1 on both sides and divide out by  $\lambda^2$  the resulting inequality. Then (3.2.4) follows easily by Taylor expansion and letting  $\lambda$  tend to 0.

Now, let's give some examples.

**Dobrushin's uniqueness regime.** The following theorem ensures that a Gaussian concentration bound holds under Dobrushin uniqueness condition 2.3.8. It was proved in [50].

**Theorem 3.2.1** (GCB under Dobrushin's uniqueness condition).

Let  $\Phi \in \mathcal{B}_T$  and assume that the associated specification  $\gamma^\Phi$  satisfies Dobrushin's uniqueness condition 2.3.8. Then  $\mu_\Phi$  satisfies GCB( $D$ ) with  $D = (2(1 - \epsilon(\gamma^\Phi))^2)^{-1}$ .

This holds for instance for any finite-range potential  $\beta\Phi$  provided that  $\beta > 0$  is small enough. As a basic example, we mention the nearest-neighbor ferromagnetic Ising model without magnetic field at sufficiently high temperature which is such that there exists  $\bar{\beta} < \beta_c$  such that for all  $\beta < \bar{\beta}$ , one has, for all  $K \in \mathcal{L}$  and for all  $u > 0$

$$\mu_{\beta\Phi}\{\omega \in \Omega : K(\omega) \geq \mathbb{E}_{\mu_{\beta\Phi}}[K] + u\} \leq \exp\left(-\frac{(1 - \tanh \beta J)^2 u^2}{2\|\bar{\delta}(K)\|_2^2}\right).$$

Note that a Gaussian concentration bound does not only hold at high-temperature regime but it can also hold at sufficiently low-temperature regime with sufficiently large external magnetic field (see 2.3 and [14] for more details and examples). We will prove in chapter 6 that, in *dimension two*, we have a Gaussian concentration bound for all  $\beta < \beta_c$ .

**Remark 3.3** (The 2D Ising model at  $\beta = \beta_c$ ).

It is known (see [[33] p. 172]) that for  $\beta = \beta_c = \log(1 + \sqrt{2})/2$ , there is a unique Gibbs measure  $\mu_{\beta_c}$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{\Lambda_n} \text{Var}_{\mu_{\beta_c}} \left( \sum_{x \in \Lambda_n} \omega_x \right) = +\infty. \quad (3.2.5)$$

Therefore,  $\mu_{\beta_c}$  cannot satisfy a Gaussian concentration bound because, by (3.2.4), one would have

$$\limsup_{n \rightarrow \infty} \frac{1}{\Lambda_n} \text{Var}_{\mu_{\beta_c}} \left( \sum_{x \in \Lambda_n} \omega_x \right) \leq 8D \quad (3.2.6)$$

which contradicts (3.2.5). To obtain (3.2.6), apply (3.2.4) to

$$K(\omega) = \sum_{x \in \Lambda_n} \pi_{\{0\}}(T_x \omega),$$

where  $\pi_{\{0\}}(\omega) = \omega_0$ . Then use Lemma 4.1.1 to get

$$\text{Var}_{\mu_{\beta_c}} \left( \sum_{x \in \Lambda_n} \omega_x \right) \leq 8D|\Lambda_n|$$

for all  $n$ . Therefore, the Gibbs measure for the Ising model in dimension 2 at critical temperature does not satisfy a Gaussian concentration bound. It does not even satisfy the variance inequality (3.2.4).

Now, we give some applications and refer the reader to [14] for more details and more applications.

**Empirical magnetization** The magnetization is of particular interest in the context of Gibbs measures. Let  $\Lambda \in \mathbb{Z}^d$ , we define the empirical magnetization in  $\Lambda$  by

$$M_\Lambda(\omega) = \sum_{x \in \Lambda} \omega_x.$$

We have the following theorem.

**Theorem 3.2.2** ([14]).

Let  $\Phi \in \mathcal{B}_T$  and assume that the associated specification  $\gamma^\Phi$  satisfies Dobrushin's uniqueness condition 2.3.8. Then, for all  $\Lambda \in \mathbb{Z}^d$ , we have

$$\mu_\Phi \left( \omega \in \Omega : \left| \frac{M_\Lambda(\omega)}{|\Lambda|} - \mathbb{E}_{\mu_\Phi}[\omega_0] \right| \geq u \right) \leq 2 \exp \left( -\frac{(1 - \mathfrak{c}(\gamma^\Phi))^2}{8} |\Lambda| u^2 \right).$$

**Speed of convergence of the empirical measure.** The aim is to quantify the speed of convergence of empirical measures in 2.2.5. We endow  $\mathcal{M}(\Omega)$  with the Kantorovich distance  $d_K$  defined by

$$d_K(\mu_1, \mu_2) = \sup \{ \mathbb{E}_{\mu_1}[G] - \mathbb{E}_{\mu_2}[G] : G : \Omega \rightarrow \mathbb{R} \text{ such that } G \text{ 1-Lipshitz} \}.$$

A function  $G : \Omega \rightarrow \mathbb{R}$  is 1-Lipschitz if  $|G(\omega) - G(\omega')| \leq d(\omega, \omega')$  where the distance  $d(\cdot, \cdot)$  is defined in 2.1.1.

**Theorem 3.2.3** ([14]).

Let  $\Phi \in \mathcal{B}_T$  and assume that the associated specification  $\gamma^\Phi$  satisfies Dobrushin's uniqueness condition 2.3.8. Denote by  $\mu_\Phi$  the corresponding Gibbs measure. Then

$$\begin{aligned} \mu_\Phi \left( \omega \in \Omega : |d_K(\mathcal{E}_n(\omega), \mu_\Phi) - \mathbb{E}_{\mu_\Phi} [d_K(\mathcal{E}_n(\cdot), \mu_\Phi)]| \geq u \right) \\ \leq 2 \exp \left( -\frac{(1 - \mathfrak{c}(\gamma^\Phi))^2}{2c_d} |\Lambda_n| u^2 \right) \end{aligned}$$

for all  $n \geq 1$  and for all  $u > 0$ , where  $c_d$  is a constant only depending on  $d$ .

**Remark 3.4.** For convenience, we decided to deal with empirical measures on "cubes" but this theorem remains valid for all  $\Lambda \in \mathbb{Z}^d$ .

**Bounding  $\bar{d}$ -distance by relative entropy.** Given  $n \in \mathbb{N}$ , we define the non normalized Hamming distance between  $\omega$  and  $\omega'$  in  $\Omega_n$  by

$$\bar{d}_n(\omega, \omega') = \sum_{x \in \Lambda_n} \mathbb{1}_{\omega_x \neq \omega'_x}.$$

Let  $\mu, \nu \in \mathcal{M}_T(\Omega)$  and define the  $\bar{d}$ -distance as follows.

$$\bar{d}(\mu, \nu) = \inf_{\mathbb{P}_n \in \mathcal{J}_T(\mu_n, \nu_n)} \int_{\Omega_n} \int_{\Omega_n} \bar{d}_n(\omega, \omega') d\mathbb{P}_n(\omega, \omega')$$

where  $\mathcal{J}_T(\mu_n, \nu_n)$  denotes the set of all shift-invariant couplings of  $\mu_n$  and  $\nu_n$ . By [73], it follows that  $\bar{d}(\mu_n, \nu_n)$  normalized by  $(2n + 1)^d$  converges to a limit that we denote by  $\bar{d}(\mu, \nu)$ . This limit defines a distance on the set on  $\mathcal{M}_T(\Omega)$ .

**Theorem 3.2.4** ([14]).

Let  $\Phi \in \mathcal{B}_T$  and assume that the associated specification  $\gamma^\Phi$  satisfies Dobrushin's uniqueness condition 2.3.8. Then for every shift-invariant probability measure  $\nu$

$$\bar{d}(\mu_\Phi, \nu) \leq \frac{\sqrt{2}}{1 - \epsilon(\gamma^\Phi)} \sqrt{h(\nu|\mu_\Phi)}.$$

**3.2.2 Moment concentration bounds**

For the Ising model ( $d \geq 2$ ) without external magnetic field, at sufficiently low temperature, one cannot expect Gaussian concentration inequalities because it would contradict the “surface-order” large deviation for the magnetization. However, the authors of [13] showed that one can control all moments and deduce that moment concentration bounds hold. We give the general form of such bound.

**Definition 3.2.2** (Moment concentration bound).

Let  $\mu$  be a probability measure on  $(\Omega, \mathfrak{B})$ . Given  $p \in \mathbb{N}$ , we say that  $\mu$  satisfies a moment concentration bound of order  $2p$  with constant  $C_{2p} = C_{2p}(\mu) > 0$  (abbreviated  $MCB(2p, C_{2p})$ ) if, for all function  $K \in \mathcal{L}$ , we have

$$\mathbb{E}_\mu[(K - \mathbb{E}_\mu[K])^{2p}] \leq C_{2p} \|\underline{\delta}(K)\|_2^2. \quad (3.2.7)$$

In the same vein as the proposition 3.2.1, we have for all  $u > 0$

$$\mu\{\omega \in \Omega : |K(\omega) - \mathbb{E}_\mu[K]| \geq u\} \leq \frac{C_{2p} \|\underline{\delta}(K)\|_2^{2p}}{u^{2p}}. \quad (3.2.8)$$

**Theorem 3.2.5** ([13, 14]).

Let  $\mu_{\beta\Phi}^+$  be the plus-phase of the low-temperature Ising model. There exists  $\bar{\beta} > \beta_c$ , such that for each  $\beta > \bar{\beta}$ , there exists a positive sequence  $(C_{2p}(\beta))_{p \in \mathbb{N}}$  such that  $\mu_{\beta\Phi}^+$  satisfies  $MCB(2p, C_{2p}(\beta))$  for all  $p \in \mathbb{N}$ .

For this model, one may wonder whether the previous theorem implies a stronger statement like a Gaussian concentration bound. In fact, one cannot infer a Gaussian concentration bound because  $C_{2p}$  is of the form  $p^{2p}G^p$  for some constant  $G > 0$  depending on  $K$  but not of  $p$  (see Theorem 3 in [13, 14]). For some  $C_{2p}$  which does not “grow too fast”, Gaussian concentration is equivalent to moment concentration for all  $p \in \mathbb{N}$  (see [10]).

**3.2.3 Stretched-exponential concentration bound**

For the Ising model at sufficiently low temperature, it turns out that the control of all moments leads to a stretched-exponential concentration bound which we define as follows. Let  $0 < \rho < 1$  and define the Young function by  $M_\rho : \mathbb{R} \rightarrow \mathbb{R}^+$  such that for all  $x \in \mathbb{R}$

$$M_\rho(x) := e^{(|x|+h_\rho)^\rho} - e^{h_\rho^\rho}$$

where  $h_\rho = \left(\frac{1-\rho}{\rho}\right)^{\frac{1}{\rho}}$ . We also define the Luxemburg norm with respect to  $M_\rho$  of a real-valued random variable  $K$  as

$$\|K\|_{M_\rho} = \inf \left\{ \lambda > 0 : \mathbb{E} \left[ M_\rho \left( \frac{K}{\lambda} \right) \right] \leq 1 \right\}.$$

Let us mention that the set of all functions  $K : \Omega \rightarrow \mathbb{R}$  such that  $\|K\|_{M_\rho} < +\infty$  is called a Orlicz space. These spaces generalize  $L_\mu^p$  spaces.

**Definition 3.2.3** (Stretched-exponential concentration bound).

Let  $\mu$  be a probability measure on  $(\Omega, \mathfrak{B})$ . Given  $p \in \mathbb{N}$ , we say that  $\mu$  satisfies a stretched-exponential concentration if there exists  $\rho = \rho(\mu) \in (0, 1)$  and a constant  $D_\rho > 0$  (abbreviated  $SECB(\rho, D_\rho)$ ) such that, for all function  $K \in \mathcal{L}$ , we have

$$\|K - \mathbb{E}_\mu[K]\|_{M_\rho} \leq D_\rho \|\underline{\delta}(K)\|_2. \quad (3.2.9)$$

This general definition leads to the following proposition.

**Proposition 3.2.3** ([13]).

If a probability measure  $\mu$  on  $(\Omega, \mathfrak{B})$  satisfies  $SECB(\rho, D_\rho)$  then there exists  $C_\rho$  such that for all function  $K \in \mathcal{L}$  and for all  $u > 0$ ,

$$\mu\{\omega \in \Omega : |K(\omega) - \mathbb{E}_\mu[K]| \geq u\} \leq 4 \exp\left(-\frac{D_\rho u^\rho}{\|\underline{\delta}(K)\|_2^\rho}\right). \quad (3.2.10)$$

For the plus-phase of the low temperature Ising model, we mention the theorem in [13, 14].

**Theorem 3.2.6** ([13, 14]).

Let  $\mu_{\beta\Phi}^+$  be the plus-phase of the low-temperature Ising model. There exists  $\bar{\beta} > \beta_c$ , such that for each  $\beta > \bar{\beta}$ ,  $\mu_{\beta\Phi}^+$  satisfies  $SECB(\rho, D_\rho)$ .

It is unlikely that the dependence of  $u$  is optimal. The constants  $\rho$  and  $D_\rho$  appearing in this theorem are not explicit and depend on the dimension  $d$  and the temperature. As a matter of fact, let consider  $K : \Omega \rightarrow \mathbb{R}$  as the empirical magnetization in  $\Lambda$ , then one has the following.

**Theorem 3.2.7** ([13, 14]).

Let  $\mu_{\beta\Phi}^+$  be the plus-phase of the low-temperature Ising model. There exists  $\bar{\beta} > \beta_c$ , such that for each  $\beta > \bar{\beta}$ , there exist  $\rho = \rho(\beta) \in (0, 1)$  and a constant  $D_\rho$  such that for all  $\Lambda \in Z^d$ , we have

$$\mu_{\beta\Phi}^+ \left( \omega \in \Omega : \left| \frac{M_\Lambda(\omega)}{|\Lambda|} - \mathbb{E}_{\mu_{\beta\Phi}^+}[\omega_0] \right| \geq u \right) \leq 4 \exp\left(-\frac{D_\rho}{2^\rho} |\Lambda|^{\frac{\rho}{2}} u^\rho\right).$$

At low temperature, one has ‘‘surface-order’’ large deviations (see [76]). In particular, let  $a, b$  such that

$$-\mathbb{E}_{\mu_{\beta\Phi}^+}[\omega_0] < a < b < \mathbb{E}_{\mu_{\beta\Phi}^+}[\omega_0],$$



then the probability (under  $\mu_\beta^+$ ) that  $M_{\Lambda_n}/|\Lambda_n|$  falls into an interval  $[a, b]$  is exponentially small in  $(2n + 1)^{d-1}$ , as  $n$  goes to infinity. This result is actually sharper than the stretched-exponential bound (since  $\rho d/2 < d-1$  for all  $d \geq 2$ ) which is valid in any finite volume and for a large class of functions. Later, we will investigate numerically this behavior.

### 3.3 Known results for stochastic chains of unbounded memory (SCUMs)

In view of Chapter 7, we give recent results on concentration inequalities for SCUMs which are a natural generalization of Markov chains by considering the whole “past”. In the literature on stochastic processes, SCUMs are also known as “chains with complete connections”. Such one-dimensional systems are simultaneously studied as discrete-time processes in symbolic dynamics under the name of “g-measures” and as Gibbs measures on the lattice  $\mathbb{N}$  or  $\mathbb{Z}$  associated to a family of specifications. For further details, we refer the reader to R. Fernández and G. Maillard [36] and S. Berghout, R. Fernández and E. Verbitskiy [6] where they study the similarities and the differences between these different objects. Let  $\Omega = S^{\mathbb{N}}$  where  $S$  is a finite set. We consider the following class of functions. Let  $n \in \mathbb{N}$  and  $K : S^n \rightarrow \mathbb{R}$ , and let

$$\delta_j(K) = \sup_{a_j=b_j, \forall j \neq i} \{|K(a_0, \dots, a_i, \dots, a_{n-1}) - K(b_0, \dots, b_i, \dots, b_{n-1})|\}. \quad (3.3.1)$$

Now let  $K : S^n \rightarrow \mathbb{R}$  and

$$\|\underline{\delta}(K)\|_2^2 := \sum_{j=-\infty}^{+\infty} \delta_j(K)^2 \quad (3.3.2)$$

where  $\underline{\delta}(K)$  is the column vector of size  $n$  whose  $j$ -th coordinate is  $\delta_j(K)$ . We stress that this class of functions is analogous to the  $d$ -dimensional case.

We will say that a measure  $\mu$  on  $(\Omega, \mathfrak{B})$  satisfies a Gaussian concentration bound if it satisfies 3.2.1 for all  $n \in \mathbb{N}$  and  $K : S^n \rightarrow \mathbb{R}$ . This statistical property is derived from the control of the dependence of the past of the probability transitions quantified by either the variation or the oscillation (see [38, 16]). Here, we will only mention the variation of a kernel/potential defined as follows:

$$\text{var}_n(\phi) := \sup\{|\phi(x) - \phi(y)| : x_i = y_i, 0 \leq i \leq n-1\}.$$

In [38], the authors proved a Gaussian concentration bound for potentials with summable variations. Recently, J.R Chazottes, S. Gallo and D. Takahashi have generalized the previous result obtaining optimal Gaussian concentration bounds for chains with complete connections on a countably infinite alphabet. Their proofs rely on maximal coupling techniques and on the tight control of the summability of its variations (see [16]).

In Chapter 7, we will obtain Gaussian concentration bounds for separately Lipschitz functions on  $\Omega^n$ . To do so, we use the spectral properties of Ruelle-Perron-Frobenius’ operator associated to the potential and, in particular its speed of convergence to the unique equilibrium state for Lipschitz observables. Of course, this result also holds for all  $K : S^n \rightarrow \mathbb{R}$ . We will deduce symbolic dynamics applications of concentration.

### 3.4 On the relation between concentration and large deviations for equilibrium states

We refer to [82] for more details and references on large deviations. An interesting consequence is that it provides a statistical interpretation of relative entropy 2.3.5 (see [82]). We will briefly explain the differences between Gaussian concentration bounds and large deviations.

First, let's consider equilibrium states for  $\Phi$ ,  $\mu \in \mathcal{G}_T(\Phi)$  and local observables  $K : \Omega \rightarrow \mathbb{R}$ . We define the Birkhoff sum of  $K$  as

$$S_n^K := \sum_{x \in \Lambda_n} K \circ T_x.$$

Then, by Theorem 2.2.1,  $S_n^K/|\Lambda_n|$  converges in probability to the expected value  $\int K d\mu$ . This can be reformulated as follows: If  $I$  is a closed interval of  $\mathbb{R}$  which does not contain  $\int K d\mu$  then

$$\lim_{n \rightarrow \infty} \mu \left\{ \omega \in \Omega : \frac{S_n^K(\omega)}{|\Lambda_n|} \in I \right\} = 0.$$

The aim of large deviations theory is to strengthen the ergodic theorem in the sense that it gives asymptotically the precise rate of convergence of this probability on the logarithmic scale. Let  $\omega \in \Omega$  and  $K \in \mathcal{L}$ , then the large deviation principle states that for all closed interval  $I$  of  $\mathbb{R}$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \mu \left\{ \omega \in \Omega : \frac{S_n^K(\omega)}{|\Lambda_n|} \in I \right\} \\ \leq - \inf \left\{ h(\nu|\mu) : \nu \in \mathcal{M}_T(\Omega) \text{ with } \int K d\nu \in I \right\} \end{aligned}$$

and for all open intervals  $I$  of  $\mathbb{R}$

$$\begin{aligned} - \inf \left\{ h(\nu|\mu) : \nu \in \mathcal{M}_T(\Omega) \text{ with } \int K d\nu \in I \right\} \\ \leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \mu \left\{ \omega \in \Omega : \frac{S_n^K(\omega)}{|\Lambda_n|} \in I \right\}. \end{aligned}$$

We can formulate a more abstract large deviation principle on the level of empirical measures  $(\mathcal{E}_n)_{n \in \mathbb{N}}$  (2.2.4). For all subsets  $A \subset \mathcal{M}(\Omega)$  which do not contain  $\mu$ , if  $A$  is a closed set then

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \mu \{ \mathcal{E}_n \in A \} \leq - \inf \{ h(\nu|\mu) : \nu \in A \cap \mathcal{M}_T(\Omega) \} \quad (3.4.1)$$

and if  $A$  is an open set then

$$- \inf \{ h(\nu|\mu) : \nu \in A \cap \mathcal{M}_T(\Omega) \} \leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \mu \{ \mathcal{E}_n \in A \}. \quad (3.4.2)$$

If we consider  $\mu$  as a product measure, the events  $\{ \mathcal{E}_n \in A \}$  are sufficient to describe completely the measure. In the case of Gibbs measures, one needs observables which depends on a large number of sites to describe the dependence in between. Then, the large deviation principle describes the relative entropy in terms of rate functions, *i.e.*, the probability that a configuration  $\omega$  distributed according to  $\mu$  looks like a typical configuration from  $\nu$  in  $\Lambda_n$  decays exponentially fast with  $n$  with rate  $h(\nu|\mu)$ .

Now, we start our comparison with concentration by considering particular events of the form

$$\left\{ \omega \in \Omega : \frac{S_n^K(\omega)}{|\Lambda_n|} - \int K d\mu \geq u \right\}$$

where  $u > 0$ . For the sake of simplicity, let's assume that  $\int K d\mu = 0$ . By Theorem 3.1.1, we have

$$\mu \left\{ \omega \in \Omega : \lambda S_n^K(\omega) \geq \lambda u |\Lambda_n| \right\} \leq e^{-|\Lambda_n| \lambda u} \mathbb{E}_\mu \left[ e^{\lambda S_n^K(\omega)} \right]$$

or, equivalently

$$\frac{1}{|\Lambda_n|} \log \mu \left\{ \omega \in \Omega : \lambda S_n^K(\omega) \geq \lambda u |\Lambda_n| \right\} \leq -\lambda u + \frac{1}{|\Lambda_n|} \log \mathbb{E}_\mu \left[ e^{\lambda S_n^K(\omega)} \right]$$

where  $\lambda, u > 0$ . Hence, when the Gaussian concentration bound (3.2.2) holds for the Birkhoff sum of  $K$ , we obtain

$$\begin{aligned} \frac{1}{|\Lambda_n|} \log \mu \left\{ \omega \in \Omega : \lambda S_n^K(\omega) \geq \lambda u |\Lambda_n| \right\} &\leq -\sup_{\lambda > 0} (\lambda u - D \lambda^2 \|\delta(K)\|_2^2) \\ &= \frac{u^2}{4D \|\delta(K)\|_2^2}. \end{aligned}$$

Notice that the upper-bound only depends on  $K$  but not on  $n$ . We observe that large deviations and concentration inequalities play complementary roles. The first one gives an asymptotic exact result but only deals with Birkhoff sums. Whereas, the second one provides a non-asymptotic upper-bound for more general functions which can be in particular non-linear and implicitly defined. For concentration inequalities the constants are typically not exact, whereas large-deviation rate functions are exact. In that sense large deviations are stronger but the non-asymptotic character and the generality of functions  $K$  makes concentration inequalities stronger.

## Chapter 4

# Gaussian concentration and uniqueness of equilibrium states in lattice systems

In this chapter we prove in two different ways that non-uniqueness of the equilibrium states for a potential prevents Gaussian concentration. This is the content of my paper [19] written in collaboration with J.-R. Chazottes, F. Redig and E. Ugalde and it is formulated in its contrapositive form as follows.

### Theorem 4.0.1.

*If an ergodic equilibrium state for a shift-invariant absolutely summable potential satisfies a Gaussian concentration bound, then it must be the unique equilibrium state for this potential.*

Let us give some examples.

**Ising model** Let's consider the nearest-neighbor ferromagnetic Ising model on  $\mathbb{Z}^d$  with zero external magnetic field (see Section 2.3). We recall that for all  $\beta > \beta_c$  there exists two distinct ergodic Gibbs measures  $\mu_{\beta\Phi}^+$  and  $\mu_{\beta\Phi}^-$ . Then, according to our main theorem, they cannot satisfy GCB. In fact, they satisfies a stretched-exponential concentration bound. For  $d = 2$ , there is an inverse temperature  $\beta_c$  such that, for all  $\beta \leq \beta_c$ ,  $\mathcal{G}_T(\beta\Phi)$  is a singleton, and, for all  $\beta > \beta_c$ ,  $\mathcal{G}_T(\beta\Phi) = \{\lambda\mu_{\beta\Phi}^+ + (1 - \lambda)\mu_{\beta\Phi}^- : \lambda \in [0, 1]\}$  (hence  $\text{ex } \mathcal{G}_T(\beta\Phi) = \{\mu_{\beta\Phi}^+, \mu_{\beta\Phi}^-\}$ ). For more details, we refer the reader to the following chapter.

For  $d = 3$ , the situation is more complicated at low temperatures. Indeed, in addition to  $\mu_{\beta\Phi}^+$  and  $\mu_{\beta\Phi}^-$ ,  $\mathcal{G}(\beta\Phi)$  also contains, for  $\beta$  large enough, a family of Gibbs measures which are not shift-invariant, the so-called Dobrushin states. In other words,  $\mathcal{G}(\beta\Phi) \setminus \mathcal{G}_T(\beta\Phi) \neq \emptyset$ . In the present chapter, we do not deal with these non-shift invariant Gibbs measures for the Ising model (and other models). Indeed, whereas for translation invariant Gibbs states we obtain here a general uniqueness result, the situation becomes much more intricate for non-translation Gibbs states, and even more for non-translation invariant potentials. In fact, Dobrushin interface states can be shown to be incompatible with GCB, using the volume large deviation bound (4.3.1). However, other more subtle scenarios of non-uniqueness combined with a unique translation invariant Gibbs measure can occur, such as in [9], and with the techniques developed here, we cannot show that GCB excludes such sce-

narios of non-uniqueness.

**Long-range Ising model** For the long-range Ising model defined in Section 2.3 the situation is the same, since there exists two distinct equilibrium states  $\mu_{\beta\Phi}^+$  and  $\mu_{\beta\Phi}^-$  in the low-temperature regime. By the above theorem, these two equilibrium states cannot satisfy a Gaussian concentration bound.

**Potts model** For the Potts model defined in Section 2.3, we recall that there exists  $0 < \beta_N < +\infty$  such that  $|\text{ex } \mathcal{G}(\beta_N \Phi)| = N+1$  and  $|\text{ex } \mathcal{G}(\beta \Phi)| = N$  when  $\beta > \beta_N$ . Then they cannot satisfy a Gaussian concentration bound.

Let us comment on the two proofs. The first way is based on ideas which were put forward in ergodic theory to study the existence of finitary codings from a finite-valued i.i.d. process to certain ergodic processes. In the context of  $g$ -measures, the authors in [38] proved with similar methods a characterization of uniqueness of the stationary chain which is a stronger statement that Gaussian concentration implies uniqueness. Two central notions turn out to be the “blowing-up property” and the “positive (lower) relative entropy property”. Without going into detail, it was proved in [65] that if a process is finitely determined then it has the blowing-up property, which in turn implies the positive relative entropy property. Here, we use the fact that the Gaussian concentration bound implies the blowing-up property, and then use part of the variational principle which says that the relative entropy of two distinct equilibrium states for the same potential is equal to zero. Hence the blowing-up property cannot hold, therefore it is not possible to have a Gaussian concentration bound. In fact, we establish an abstract result (Theorem 4.2.1) which states that if a probability measure satisfies a Gaussian concentration bound, then it has the positive lower relative entropy property. Technically speaking, we follow some methods that can be found in [65]. The passage from  $d = 1$  to  $d \geq 2$  (that is, going from processes to random measures) poses no difficulty. Although these methods are known to specialists of ergodic theory, they are probably not as well-known in the mathematical physics literature. The other way of proving the main theorem is via large deviations which is shorter and simpler than the first one, given that we have a large deviation principle at our disposal.

## 4.1 Setting

Let  $\Omega = S^{\mathbb{Z}^d}$  where  $S$  is a finite set and  $d \geq 1$ . We recall that the shift action  $(T_x, x \in \mathbb{Z}^d)$  is defined this way: for each  $x \in \mathbb{Z}^d$   $T_x : \Omega \rightarrow \Omega$  and  $(T_x \omega)_y = \omega_{y-x}$ , for all  $y \in \mathbb{Z}^d$ . We also recall what we mean by Gaussian concentration bound in this chapter. Let  $K : \Omega \rightarrow \mathbb{R}$  a continuous function and  $x \in \mathbb{Z}^d$ . We denote by

$$\delta_x(K) = \sup \{ |K(\omega) - K(\omega')| : \omega, \omega' \in \Omega \text{ differ only at site } x \}$$

the oscillation of  $K$  at  $x$ . We say that  $K : \Omega \rightarrow \mathbb{R}$  is a *local function* if there exists  $\Lambda_K \Subset \mathbb{Z}^d$  (the dependence set of  $K$ ) such that for all  $\omega, \omega', \omega'' \in \Omega$   $K(\omega_{\Lambda_K} \omega'_{\Lambda_K^c}) = K(\omega_{\Lambda_K} \omega''_{\Lambda_K^c})$ . Equivalently,  $\delta_x(K) = 0$  for all  $x \notin \Lambda_K$ . It is understood that  $\Lambda_K$  is the smallest such set. Local functions are continuous,

hence bounded since  $\Omega$  is compact. (In fact, continuous functions are obtained as uniform limits of local functions).

We write  $\underline{\delta}(K)$  for the infinite array  $(\delta_x(K), x \in \mathbb{Z}^d)$ , and let

$$\|\underline{\delta}(K)\|_2^2 := \|\underline{\delta}(K)\|_{\ell^2(\mathbb{Z}^d)}^2 = \sum_{x \in \mathbb{Z}^d} \delta_x(K)^2.$$

For a local function  $K$ , it may happen that  $\delta_x(K) = +\infty$  for some  $x$  in its dependence set, whence  $\|\underline{\delta}(K)\|_2 = +\infty$ . If  $S$  is finite, this cannot happen. Recall that

$$\mathcal{L} = \bigcup_{\Lambda \in \mathbb{Z}^d} \mathcal{L}_\Lambda \quad (4.1.1)$$

where

$$\mathcal{L}_\Lambda = \{K : \Omega \rightarrow \mathbb{R} : \delta_x(K) = 0 \forall x \notin \Lambda\},$$

and by using a simple telescoping argument, it can be checked that any  $K \in \mathcal{L}$  is bounded. We also recall what we mean by Gaussian concentration bound.

**Definition 4.1.1** (Gaussian concentration bound).

Let  $\mu$  be a probability measure on  $(\Omega, \mathfrak{B})$ . We say that it satisfies a Gaussian concentration bound if there exists  $D = D(\mu) > 0$  such that, for all functions  $K \in \mathcal{L}$ , we have

$$\mathbb{E}_\mu [\exp(K - \mathbb{E}_\mu[K])] \leq \exp(D\|\underline{\delta}(K)\|_2^2). \quad (4.1.2)$$

The next result shows that a shift-invariant probability measure which satisfies GCB must be mixing.

**Proposition 4.1.1.**

Let  $\mu$  be a shift-invariant probability measure on  $(\Omega, \mathfrak{B})$  which satisfies GCB( $D$ ). Then  $\mu$  is mixing.

*Proof.* First we introduce the following result that we will prove in the appendix.

**Lemma 4.1.1.**

Let  $f : \Omega \rightarrow \mathbb{R}$  such that  $\|\underline{\delta}(f)\|_1 := \sum_{x \in \mathbb{Z}^d} \delta_x(f) < +\infty$ . Then for any  $\Lambda \in \mathbb{Z}^d$  we have

$$\|\underline{\delta}(S_\Lambda f)\|_2^2 \leq |\Lambda| \|\underline{\delta}(f)\|_1^2.$$

By this Lemma and Proposition 3.2.1, we conclude that for every sequence  $V_n, n \in \mathbb{N}$  of finite subsets of  $\mathbb{Z}^d$  such that  $|V_n| \rightarrow \infty$  as  $n \rightarrow \infty$ , and for every local function  $f$ , we have that  $\frac{1}{|V_n|} \sum_{x \in V_n} f \circ T_x$  converges to  $\mathbb{E}_\mu(f)$  in  $\mu$ -probability as  $n \rightarrow \infty$ .

We then argue by contradiction. Assume that  $\mu$  is not mixing. Then there exist local functions  $f, g$  (without loss of generality both of  $\mu$  expectation zero) and a sequence  $x_n$ , with  $|x_n| \rightarrow \infty$  such that

$$\int f \cdot g \circ T_{x_n} d\mu$$

does not converge to zero as  $n \rightarrow \infty$ . By locality, both functions  $f, g$  are uniformly bounded, and therefore the sequence  $\int f \cdot g \circ T_{x_n} d\mu, n \in \mathbb{N}$  is a bounded sequence. Therefore, there exists a subsequence  $y_n$  such that along that subsequence

$$\lim_{n \rightarrow \infty} \int f \cdot g \circ T_{y_n} d\mu = a \neq 0. \quad (4.1.3)$$

As a consequence,

$$\lim_{n \rightarrow \infty} \int f \cdot \frac{1}{n} \sum_{k=0}^n g \circ T_{y_k} d\mu = a \neq 0.$$

However, as we saw before,  $\frac{1}{n} \sum_{k=0}^n g \circ T_{y_k}$  converges to zero in probability by GCB. Then via dominated convergence we obtain

$$\lim_{n \rightarrow \infty} \left| \int f \cdot \frac{1}{n} \sum_{k=0}^n g \circ T_{y_k} d\mu \right| = 0$$

which contradicts (4.1.3).  $\square$

As direct consequence,  $\mu$  is ergodic but we give another proof that a shift-invariant probability measure which satisfies GCB must be ergodic.

**Proposition 4.1.2.**

*Let  $\mu$  be a shift-invariant probability measure on  $(\Omega, \mathfrak{B})$  which satisfies GCB. Then  $\mu$  is ergodic.*

*Proof.* We will use the notation  $S_n f = \sum_{x \in \Lambda_n} f \circ T_x$  for a function  $f : \Omega \rightarrow \mathbb{R}$ . Let's recall that by Proposition 2.2.1 a probability measure  $\mu$  on  $(\Omega, \mathfrak{B})$  is ergodic for the shift if and only if, for any pair  $A, B \in \mathfrak{B}$ , one has

$$\lim_{n \rightarrow \infty} \frac{1}{(2n+1)^d} \sum_{x \in \Lambda_n} \mu(A \cap T_x B) = \mu(A)\mu(B). \quad (4.1.4)$$

Moreover, this equivalence is true if we take  $A, B$  as a cylinder sets. (We refer to [[44], Lemma 2.2.2 p. 30] for the proofs of these facts.) We use the notation  $\int f d\mu$  instead of  $\mathbb{E}_\mu(K)$  as it is more convenient here. We are going to prove that, for any pair  $A, B$  of cylinder sets, (3.2.3) implies

$$\lim_{n \rightarrow \infty} \int (\mathbb{1}_A - \mu(A)) \frac{S_n \mathbb{1}_B}{(2n+1)^d} d\mu = 0 \quad (4.1.5)$$

which implies at once (4.1.4) since by the shift-invariance of  $\mu$

$$\int (\mathbb{1}_A - \mu(A)) \frac{S_n \mathbb{1}_B}{(2n+1)^d} d\mu = \frac{1}{(2n+1)^d} \sum_{x \in \Lambda_n} \int \mathbb{1}_A \cdot \mathbb{1}_B \circ T_x d\mu - \mu(A)\mu(B).$$

To prove (4.1.5), we use that  $S_n \mathbb{1}_B / (2n+1)^d$  converges in probability to  $\int \mathbb{1}_B d\mu = \mu(B)$ . Indeed, using Lemma and (3.2.3), that we apply to  $F = S_n \mathbb{1}_B / (2n+1)^d$ , we have

$$\mu \left\{ \omega \in \Omega : \left| \frac{S_n \mathbb{1}_B(\omega)}{(2n+1)^d} - \mu(B) \right| \geq u \right\} \leq \exp \left( - \frac{(2n+1)^d u^2}{4D \|\underline{\delta}(\mathbb{1}_B)\|_1^2} \right)$$

for all  $u > 0$ , where  $\|\underline{\delta}(\mathbb{1}_B)\|_1$  is obviously finite. (More precisely, we apply (3.2.3) to  $K$ , then to  $-F$ , and we finally use a union bound.) Therefore, by Lebesgue's dominated convergence theorem we get

$$\lim_{n \rightarrow \infty} \int (\mathbb{1}_A - \mu(A)) \frac{S_n \mathbb{1}_B}{(2n+1)^d} d\mu = \int (\mathbb{1}_A - \mu(A)) \mu(B) d\mu = 0.$$

The proof is finished.  $\square$

## 4.2 Proof of Theorem 4.0.1

### 4.2.1 An abstract result

The first way to prove the theorem is to establish an abstract theorem which is of independent interest, and can be applied to equilibrium states. To state it, we need to define the “positive relative entropy property”.

**Definition 4.2.1.**

An ergodic measure  $\mu$  on  $\Omega$  is said to have the positive relative entropy property if  $h_*(\nu|\mu) > 0$  for any ergodic measure  $\nu \neq \mu$ .

The lower relative entropy  $h_*(\nu|\mu)$  is defined in (2.3.5). We can now state the following result. (Recall that by Proposition 4.1.2 we only need to consider ergodic measures.)

**Theorem 4.2.1** (GCB implies positive relative entropy).

Let  $\mu$  be an ergodic probability measure on  $\Omega$  which satisfies GCB. Then  $\mu$  has the positive relative entropy property.

*Proof.* We outline the proof which consists in the following three steps. We first prove that GCB implies the so-called “blowing-up” property (Section 4.2.2). Then we prove that the blowing-up property implies the “exponential rate of convergence for frequencies” (Section 4.2.3). Finally we prove that the latter implies the positive relative entropy property (Section 4.2.4).  $\square$

**Corollary 4.2.1.**

Let  $\Phi \in \mathcal{B}_T$  and assume that  $\mathcal{G}_T(\Phi) \neq \emptyset$ . If  $|\text{ex } \mathcal{G}_T(\Phi)| \geq 2$  and  $\mu \in \text{ex } \mathcal{G}_T(\Phi)$  then  $\mu$  cannot satisfy a Gaussian concentration bound.

*Proof.* Fix an arbitrary  $\mu \in \text{ex } \mathcal{G}_T(\Phi)$ . By assumption there exists  $\mu' \in \text{ex } \mathcal{G}_T(\Phi)$  such that  $\mu \neq \mu'$ . But, by Theorem 2.3.1, we have  $h(\mu'|\Phi) = h_*(\mu'|\mu) = 0$ , whence  $\mu$  does not have the positive relative entropy property. Therefore, according to Theorem 4.2.1,  $\mu$  cannot satisfy GCB. This proves that none of the elements of  $\text{ex } \mathcal{G}_T(\Phi)$  can satisfy GCB.  $\square$



## 4.2.2 GCB implies the blowing-up property

Let  $\Lambda \in \mathbb{Z}^d$ . We define the (non-normalized) Hamming distance between two configurations  $\omega$  and  $\eta$  in  $S^\Lambda$  by

$$\bar{d}_\Lambda(\omega, \eta) = \sum_{x \in \Lambda} \mathbb{1}_{\{\omega_x \neq \eta_x\}} \in \{0, 1, \dots, |\Lambda|\}.$$

So we count at how many sites the configurations  $\omega$  and  $\eta$  in  $S^\Lambda$  differ. We can also see  $\bar{d}_\Lambda$  as a local function on  $\Omega \times \Omega$ . Given  $C \subset S^\Lambda$  define

$$\bar{d}_\Lambda(\omega, C) = \inf_{\omega' \in C} \bar{d}_\Lambda(\omega, \omega').$$

Given  $\varepsilon \in [0, 1]$ , define the  $\varepsilon$ -neighborhood (or the  $\varepsilon$ -blow-up) of  $C$  as

$$\langle C \rangle_\varepsilon = \{\omega \in S^\Lambda : \bar{d}_\Lambda(\omega, C) < \varepsilon |\Lambda|\} \subset S^\Lambda.$$

Recall that if  $C \subset S^\Lambda$ ,  $[C] = \{\omega \in \Omega : \pi_\Lambda(\omega) \in C\}$  where  $\pi_\Lambda$  is the projection from  $\Omega$  onto  $S^\Lambda$ . We now define the blowing-up property.

**Definition 4.2.2** (Blowing-up property).

An ergodic probability measure  $\mu$  on  $(\Omega, \mathfrak{B})$  has the blowing-up property if given  $\varepsilon > 0$  there is a  $\delta > 0$  and  $N \in \mathbb{N}$  such that if  $n \geq N$  and  $C \subset S^{\Lambda_n}$  then

$$\mu([C]) \geq e^{-(2n+1)d\delta} \quad \text{implies} \quad \mu(\langle C \rangle_\varepsilon) \geq 1 - \varepsilon.$$

Obviously, the blowing-up property can be formulated in terms of finite subsets of  $\mathbb{Z}^d$  of arbitrary shape, instead of cubes, but we will not need this generalization. The blowing-up property roughly says that any collection of configurations on large finite box which has a total measure which is not too exponentially small is such that most configurations are close to this collection in the Hamming distance. We have the following result.

**Proposition 4.2.1.**

Let  $\Lambda \in \mathbb{Z}^d$  and  $C \subset S^\Lambda$ . Suppose that  $\mu$  is a probability measure which satisfies GCB( $D$ ) and such that  $\mu([C]) > 0$ . Then, we have

$$\mu(\langle C \rangle_\varepsilon) \geq 1 - \exp \left[ -\frac{|\Lambda|}{4D} \left( \varepsilon - \frac{2\sqrt{D \log(\mu(C)^{-1})}}{\sqrt{|\Lambda|}} \right)^2 \right] \quad (4.2.1)$$

whenever  $\varepsilon > \frac{2\sqrt{D \log(\mu(C)^{-1})}}{\sqrt{|\Lambda|}}$ . In particular,  $\mu$  satisfies the blowing-up property.

Note that we do not require  $\mu$  to be shift-invariant. As already mentioned, if  $\mu$  is taken shift-invariant then it must be ergodic, this is enforced by the Gaussian concentration bound.

*Proof.* Consider the local function  $K(\omega) = \bar{d}_\Lambda(\omega, C)$ . One easily checks that  $\delta_x(K) \leq 1$  for all  $x \in \Lambda$ . Applying (3.2.3) gives

$$\mu\{\omega \in \Omega, : K(\omega) \geq u + \mathbb{E}_\mu(K)\} \leq \exp \left( -\frac{u^2}{4D|\Lambda|} \right) \quad (4.2.2)$$

for all  $u > 0$ . We now estimate  $\mathbb{E}_\mu(K)$  from above. Applying (3.2.1) to  $-\lambda K$ , for some  $\lambda > 0$  to be chosen later on, we get

$$\exp(\lambda \mathbb{E}_\mu(K)) \mathbb{E}_\mu(\exp(-\lambda K)) \leq \exp(D\lambda^2|\Lambda|)$$

Observe that by definition of  $K$  we have

$$\mathbb{E}_\mu(\exp(-\lambda K)) \geq \mathbb{E}_\mu(\mathbb{1}_C \exp(-\lambda K)) = \mu([C])$$

Combining the two previous inequalities and taking the logarithm gives

$$\mathbb{E}_\mu(K) \leq \inf_{\lambda > 0} \left\{ D\lambda|\Lambda| + \frac{1}{\lambda} \ln(\mu([C])^{-1}) \right\}$$

which gives the following estimate by taking the value of  $\lambda$  minimizing the right-hand side

$$\mathbb{E}_\mu(K) \leq 2\sqrt{D|\Lambda| \ln(\mu([C])^{-1})} =: v_0.$$

Therefore, inequality (4.2.2) implies that

$$\mu\{\omega \in \Omega, : K(\omega) \geq v\} \leq \exp\left(-\frac{(v - v_0)^2}{4D|\Lambda|}\right)$$

for all  $v > v_0$ . To finish the proof of (4.2.1), take  $v = \varepsilon|\Lambda|$  and observe that, by definition of  $K$ ,  $\mu\{\omega \in \Omega : K(\omega) \geq v\} = \mu(\langle C \rangle_\varepsilon^c)$ .

We now prove that  $\mu$  satisfies the blowing-up property. Let  $\varepsilon > 0$ , and take  $C$  such that  $\mu([C]) \geq e^{-(2n+1)^d \delta}$  for some  $\delta$  to be chosen later on, subject to the condition  $\varepsilon > 2\sqrt{D\delta}$ . We now apply (4.2.1) with  $\Lambda = \Lambda_n$  to get

$$\mu(\langle C \rangle_\varepsilon) \geq 1 - \exp\left(-\frac{(2n+1)^d}{4D} (\varepsilon - 2\sqrt{D\delta})^2\right).$$

Taking  $\delta = \varepsilon^2/(9D)$  yields

$$\mu(\langle C \rangle_\varepsilon) \geq 1 - \varepsilon$$

for all  $n \geq N := \lfloor (36D\varepsilon^{-2} \log \varepsilon^{-1})^{1/d} \rfloor / 2$ . □

### 4.2.3 Blowing-up implies exponential rate for frequencies

Given  $\omega \in \Omega$ ,  $n > k \geq 0$ , and a pattern  $p_k \in S^{\Lambda_k}$ , let

$$f_{n,k}(\omega; p_k) = \frac{|\{x \in \Lambda_{n-k} : (T_x \omega)_{\Lambda_k} = p_k\}|}{(2(n-k)+1)^d}. \quad (4.2.3)$$

In words, this is just the frequency of occurrence of the pattern  $p_k$  if we look at the configuration  $\omega$  restricted to the cube  $\Lambda_n$ . Let  $\mu$  be an ergodic probability measure. By the multidimensional ergodic theorem, for  $\mu$ -almost every  $\omega$ , we have

$$\lim_{n \rightarrow \infty} f_{n,k}(\omega; p_k) = \mu([p_k]).$$

Given two probability measures  $\mu$  and  $\nu$ , we denote by  $\nu_n$  (resp.  $\mu_n$ ) the probability measure induced by  $\nu$  (resp. by  $\mu$ ) on  $S^{\Lambda_n}$  by projection. The total variation distance between  $\mu_k$  and  $\nu_k$  is define by

$$\|\mu_k - \nu_k\|_{TV} = \frac{1}{2} \sum_{p_k \in S^{\Lambda_k}} |\mu([p_k]) - \nu([p_k])|.$$

We can now define the property of exponential rate of convergence for frequencies.

**Definition 4.2.3** (exponential rate of convergence property for frequencies).

An ergodic probability measure on  $(\Omega, \mathfrak{B})$  has the exponential rate of convergence property for frequencies if, given  $k$  and  $\varepsilon > 0$ , there is a  $\delta > 0$  and  $N$  such that

$$\mu(\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega; \cdot) - \mu_k\|_{TV} \geq \varepsilon) \leq e^{-(2n+1)^d \delta}, \forall n \geq N. \quad (4.2.4)$$

We now prove that this property is implied by the blowing-up property.

**Proposition 4.2.2.**

Let  $\mu$  be an ergodic probability measure on  $(\Omega, \mathfrak{B})$ . Then if  $\mu$  has the blowing-up property, it has the exponential rate of convergence property for frequencies.

*Proof.* We adapt a proof of [65] to our setting. Take  $\varepsilon > 0$  and  $k \geq 0$ , and for any  $n \geq 0$  let

$$\mathcal{B}(n, k, \varepsilon) := \{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega; \cdot) - \mu_k\|_{TV} \geq \varepsilon\}.$$

Note that we can naturally identify this subset of  $\Omega$  with a subset of  $S^{\Lambda_n}$ . We have the following lemma whose proof is given in Appendix.

**Lemma 4.2.1.**

Let  $\varepsilon$  and  $k \geq 0$ , and define

$$\rho = \frac{2\varepsilon}{5(2k+1)^d}.$$

There exists  $\tilde{N} > k$  such that, if  $n \geq \tilde{N}$  and  $\bar{d}_{\Lambda_n}(\omega, \eta) \leq \rho(2(n-k)+1)^d$ , then

$$\|\mathfrak{f}_{n,k}(\omega; \cdot) - \mathfrak{f}_{n,k}(\eta; \cdot)\|_{TV} \leq \frac{\varepsilon}{2}.$$

The lemma implies

$$\langle \mathcal{B}(n, k, \varepsilon) \rangle_\rho \subset \mathcal{B}(n, k, \frac{\varepsilon}{2}). \quad (4.2.5)$$

Now, by the blowing-up property (Definition 4.2.2), there is a  $\delta > 0$  and  $N$  such that, if  $n \geq N$  and  $C \subset S^{\Lambda_n}$  is so that  $\mu([C]) \geq e^{-(2n+1)^d \delta}$ , then  $\mu(\langle C \rangle_\rho) \geq 1 - \rho \geq 1 - \varepsilon$ . Therefore, if  $n \geq N$  and if we suppose that  $\mu(\mathcal{B}(n, k, \varepsilon)) \geq e^{-(2n+1)^d \delta}$  then we get

$$\mu(\langle \mathcal{B}(n, k, \varepsilon) \rangle_\rho) \geq 1 - \varepsilon$$

This bound together with the inclusion (4.2.5) implies for all  $n \geq \max(N, \tilde{N})$

$$\mu\left(\mathcal{B}\left(n, k, \frac{\varepsilon}{2}\right)\right) \geq 1 - \varepsilon.$$

But this contradicts the multidimensional ergodic theorem which ensures that, for each  $\varepsilon > 0$  and each  $k$ ,

$$\lim_{n \rightarrow \infty} \mu\left(\mathcal{B}\left(n, k, \frac{\varepsilon}{2}\right)\right) = 0.$$

Indeed, we have

$$\mu\left(\mathcal{B}\left(n, k, \frac{\varepsilon}{2}\right)\right) \leq \sum_{p_k \in S^{\Lambda_k}} \mu\left(\omega \in \Omega : |\mathfrak{f}_{n,k}(\omega; p_k) - \mu([p_k])| \geq \frac{\varepsilon}{|S|^{(2k+1)^d}}\right).$$

where each term on the right-hand side goes to zero by the multidimensional ergodic theorem.  $\square$

#### 4.2.4 Exponential rate for frequencies implies positive relative entropy property

We now have the following proposition

**Proposition 4.2.3.**

*Let  $\mu$  be an ergodic measure on  $(\Omega, \mathfrak{B})$ . If it has the exponential rate of convergence property for frequencies, then, given  $\varepsilon > 0$  and  $k$ , there is a  $\delta > 0$  such that, if  $\nu$  is an ergodic measure such that  $h_*(\nu|\mu) < \delta$  then  $\|\nu_k - \mu_k\|_{TV} < \varepsilon$ .*

*Proof.* We adapt a proof of [65]. Let  $\rho$  be a positive number that we will specify later on, and suppose that

$$\frac{H_n(\nu|\mu)}{(2n+1)^d} < \frac{\rho^2}{2}.$$

If  $n$  is large enough, we have by Markov inequality and lemma 4.4.1 (see appendix below) that there exists a set  $G_n \subset S^{\Lambda_n}$  such that

$$\nu(G_n) > 1 - \rho \tag{4.2.6}$$

and

$$e^{-(2n+1)^d \rho} \nu([\omega_{\Lambda_n}]) \leq \mu([\omega_{\Lambda_n}]) \leq e^{(2n+1)^d \rho} \nu([\omega_{\Lambda_n}]), \quad \omega \in G_n. \tag{4.2.7}$$

Now, for each fixed  $k$ , and for  $n$  large enough, the multidimensional ergodic theorem applied to  $\nu$  tells us that there is a set  $\tilde{G}_n \subset G_n$  such that

$$\nu(\tilde{G}_n) > 1 - 2\rho \tag{4.2.8}$$

$$\|\mathfrak{f}_{n,k}(\omega; \cdot) - \nu_k\|_{TV} < \rho, \quad \omega \in \tilde{G}_n. \tag{4.2.9}$$

If  $\mu$  has the exponential rate of convergence property for frequencies then

$$\mu\{\omega \in \Omega; \|\mathfrak{f}_{n,k}(\omega; \cdot) - \nu_k\|_{TV} \geq \rho\} \geq e^{-(2n+1)^d \tau}$$

for some  $\tau > 0$  and all  $n$  sufficiently large. Now condition (4.2.7) and (4.2.8) imply that  $\tilde{G}_n$

$$\mu(\tilde{G}_n) > (1 - 2\rho) e^{-(2n+1)^d \rho}.$$

Therefore, if  $\rho$  is small enough and  $n$  large enough, there exists an  $\omega \in \tilde{G}_n$  such that  $\|\mathfrak{f}_{n,k}(\omega; \cdot) - \mu_k\|_{TV} < \rho$ . Indeed, it is enough to check that  $\mu(\tilde{G}_n) > \mu\{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega; \cdot) - \mu_k\|_{TV} \geq \rho\}$  which implies that

$$\tilde{G}_n \cap \{\omega \in \Omega : \|\mathfrak{f}_{n,k}(\omega; \cdot) - \mu_k\|_{TV} \geq \rho\}$$

has strictly positive  $\mu$ -measure. Since (4.2.8) holds for the same  $\omega$ , we thus arrive by the triangle inequality at the estimate  $\|\nu_k - \mu_k\|_{TV} < 2\rho$ . Therefore, we proved that, given  $\varepsilon$  and  $k$ , if  $\nu$  is ergodic such that  $h_*(\nu|\mu) < \varepsilon^2/8$ , then  $\|\nu_k - \mu_k\|_{TV} < \varepsilon$ , which ends the proof.  $\square$

Now, we can state the main proposition of this section.

**Proposition 4.2.4.**

*If  $\mu$  has the exponential rate of convergence property for frequencies then it has the positive relative entropy property.*

*Proof.* Suppose that  $\mu$  has not the positive relative entropy property, but satisfies the exponential rate of convergence property for frequencies. We will obtain a contradiction. By assumption there is an ergodic measure  $\hat{\nu} \neq \mu$  such that  $h_*(\hat{\nu}|\mu) = 0$ , and there is a  $\hat{k}$  and a  $\hat{\varepsilon} > 0$  such that  $\|\hat{\mu}_{\hat{k}} - \mu_{\hat{k}}\|_{TV} \geq \hat{\varepsilon}$ . Now, we apply the Proposition 4.2.3. By the exponential rate of convergence property for frequencies, given  $\hat{\varepsilon}$  and  $\hat{k}$ , there is a  $\hat{\delta} > 0$  such that if  $\nu$  satisfies  $h_*(\nu|\mu) < \hat{\delta}$  then  $\|\nu_{\hat{k}} - \mu_{\hat{k}}\|_{TV} < \hat{\varepsilon}$ . We can take  $\nu = \hat{\nu}$ , hence we arrive at a contradiction  $\|\hat{\nu}_{\hat{k}} - \mu_{\hat{k}}\| < \hat{\varepsilon}$ .  $\square$

### 4.3 Another proof of Theorem 4.0.1

We present another proof of the theorem, based on large deviations. Suppose that  $\mu$  is an ergodic equilibrium state which satisfies a Gaussian concentration bound. Now assume that there exists  $\mu' \in \text{ex } \mathcal{G}_T(\Phi)$  such that  $\mu' \neq \mu$ . We are going to arrive at a contradiction. Notice that we can suppose that  $\mu'$  is ergodic without loss of generality, because of the ergodic decomposition [[39], Theorem 14.17, p. 298], and the fact that the map  $\nu \mapsto h(\nu|\Phi)$ ,  $\nu \in \mathcal{M}_T(\Omega)$ , is affine (which is a consequence of [[39], Theorem 15.20, p. 318] and (2.3.6)). Since  $\mu \neq \mu'$ , there exists a local function  $f : \Omega \rightarrow \mathbb{R}$  such that  $\mathbb{E}_\mu(f) \neq \mathbb{E}_{\mu'}(f)$ . Without loss of generality, assume that  $\mathbb{E}_\mu(f) < \mathbb{E}_{\mu'}(f)$ . So there exists  $\varepsilon > 0$  such that

$$\mathbb{E}_\mu(f) + \varepsilon = \mathbb{E}_{\mu'}(f).$$

We want to apply Proposition 3.2.1 to  $F = S_n f$ , where, for each  $n \geq 0$ ,  $S_n f = \sum_{x \in \Lambda_n} f \circ T_x$ . We claim that

$$\|\underline{\delta}(S_n f)\|_2^2 \leq (2n+1)^d \|\underline{\delta}(f)\|_1^2$$

where  $\|\underline{\delta}(f)\|_1 := \sum_{x \in \mathbb{Z}^d} \delta_x(f)$  (which is finite since  $K$  is local). See Lemma 4.1.1 below. Letting

$$\mathcal{E}_{n,\varepsilon} := \left\{ \omega \in \Omega : \frac{S_n f(\omega)}{(2n+1)^d} \geq \mathbb{E}_\mu(f) + \frac{\varepsilon}{3} \right\},$$

we get

$$\mu(\mathcal{E}_{n,\varepsilon}) \leq \exp\left(-\frac{(2n+1)^d \varepsilon^2}{36D \|\underline{\delta}(f)\|_1^2}\right). \quad (4.3.1)$$

Now let

$$\mathcal{E}'_{n,\varepsilon} := \left\{ \omega \in \Omega : \frac{S_n f(\omega)}{(2n+1)^d} \in \left] \mathbb{E}_{\mu'}(f) - \frac{\varepsilon}{3}, \mathbb{E}_{\mu'}(f) + \frac{\varepsilon}{3} \right[ \right\}.$$

Since

$$\mathcal{E}'_{n,\varepsilon} \subset \mathcal{E}_{n,\varepsilon}$$

we deduce that

$$\limsup_{n \rightarrow +\infty} \frac{1}{(2n+1)^d} \log \mu(\mathcal{E}'_{n,\varepsilon}) \leq -\frac{\varepsilon^2}{36D\|\underline{\delta}(f)\|_1^2} < 0. \quad (4.3.2)$$

Now we use the large deviation principle satisfied by  $\mu$  (see [[39], Section 15.5]) which implies that

$$\liminf_{n \rightarrow +\infty} \frac{1}{(2n+1)^d} \log \mu(\mathcal{E}'_{n,\varepsilon}) \geq -\inf_{u \in ]\mathbb{E}_{\mu'}(f) - \frac{\varepsilon}{3}, \mathbb{E}_{\mu'}(f) + \frac{\varepsilon}{3}[} I_f(u) \quad (4.3.3)$$

where

$$I_f(u) = \inf\{h(\nu|\mu) : \nu \in \mathcal{M}_T(\Omega), \mathbb{E}_\nu(f) = u\}.$$

The right-hand side of (4.3.3) is larger than  $-I_f(v)$  for any value of  $v$  taken in the interval, so in particular it is larger than  $-I_f(\mathbb{E}_{\mu'}(f))$ , which is equal to 0 because  $h(\mu'|\mu) = 0$  by Theorem 2.3.1. Hence, we obtain

$$\liminf_{n \rightarrow +\infty} \frac{1}{(2n+1)^d} \log \mu(\mathcal{E}'_{n,\varepsilon}) = 0$$

which contradicts (4.3.2).

## 4.4 Appendix

### 4.4.1 Proof of Lemma 4.1.1

We first recall Young's inequality for convolution. Let  $\underline{u} = (u_x)_{x \in \mathbb{Z}^d}$  and  $\underline{v} = (v_x)_{x \in \mathbb{Z}^d}$ . Formally we define their convolution  $\underline{u} * \underline{v}$  by

$$(\underline{u} * \underline{v})_x = \sum_{y \in \mathbb{Z}^d} u_{x-y} v_y, x \in \mathbb{Z}^d.$$

If  $\underline{u} \in \ell^p(\mathbb{Z}^d)$  and  $\underline{v} \in \ell^q(\mathbb{Z}^d)$ , where  $p, q \geq 1$ , then  $\underline{u} * \underline{v} \in \ell^r(\mathbb{Z}^d)$  where  $r \geq 1$  is such that  $1 + r^{-1} = p^{-1} + q^{-1}$ , then we have

$$\|\underline{u} * \underline{v}\|_{\ell^r(\mathbb{Z})} \leq \|\underline{u}\|_{\ell^p(\mathbb{Z})} \|\underline{v}\|_{\ell^q(\mathbb{Z})}.$$

Now we give the proof of Lemma 4.1.1.

*Proof.* Since  $\delta_z(S_\Lambda f) \leq \sum_{x \in \Lambda} \delta_{z-x}(f)$ , we apply Young's inequality with  $r = 2, p = 2, q = 1, u_x = \mathbb{1}_\Lambda(x)$ , and  $v_x = \delta_x(f)$  to get the desired estimate.  $\square$

### 4.4.2 Proof of Lemma 4.2.1

This lemma (in its  $d = 1$  version) is stated without proof in [68]. We fix  $\varepsilon > 0$  and  $k \geq 0$ . The frequency of a pattern  $p_k$  in  $\omega$  (see (4.2.3)) can be rewritten as

$$\mathfrak{f}_{n,k}(\omega, p_k) = \frac{1}{(2(n-k)+1)^d} \sum_{x \in \Lambda_{n-k}} \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}}.$$

By definition we have

$$\|f_{n,k}(\omega, \cdot) - f_{n,k}(\eta, \cdot)\|_{TV} = \frac{1}{2(2(n-k)+1)^d} \sum_{p_k \in S^{\Lambda_k}} \left| \sum_{x \in \Lambda_{n-k}} \left( \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}} - \mathbb{1}_{\{(T_x \eta)_{\Lambda_k} = p_k\}} \right) \right|. \quad (4.4.1)$$

Letting

$$\mathcal{I}_{\omega, \eta, n} = \{x \in \Lambda_{n-k} : (T_x \omega)_{\Lambda_k} \neq (T_x \eta)_{\Lambda_k}\}$$

we get

$$\begin{aligned} & \sum_{p_k \in S^{\Lambda_k}} \left| \sum_{x \in \Lambda_{n-k}} \left( \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}} - \mathbb{1}_{\{(T_x \eta)_{\Lambda_k} = p_k\}} \right) \right| \\ &= \sum_{p_k \in S^{\Lambda_k}} \left| \sum_{x \in \mathcal{I}_{\omega, \eta, n}^c} \left( \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}} - \mathbb{1}_{\{(T_x \eta)_{\Lambda_k} = p_k\}} \right) \right| \\ &\leq \sum_{p_k \in S^{\Lambda_k}} \sum_{x \in \mathcal{I}_{\omega, \eta, n}^c} \left| \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}} - \mathbb{1}_{\{(T_x \eta)_{\Lambda_k} = p_k\}} \right| \\ &\leq \sum_{p_k \in S^{\Lambda_k}} \sum_{x \in \mathcal{I}_{\omega, \eta, n}^c} \left( \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}} + \mathbb{1}_{\{(T_x \eta)_{\Lambda_k} = p_k\}} \right) \\ &= \sum_{x \in \mathcal{I}_{\omega, \eta, n}^c} \sum_{p_k \in S^{\Lambda_k}} \left( \mathbb{1}_{\{(T_x \omega)_{\Lambda_k} = p_k\}} + \mathbb{1}_{\{(T_x \eta)_{\Lambda_k} = p_k\}} \right) \\ &= 2 |\mathcal{I}_{\omega, \eta, n}^c|. \end{aligned}$$

Hence we obtain from (4.4.1)

$$\|f_{n,k}(\omega, \cdot) - f_{n,k}(\eta, \cdot)\|_{TV} \leq \frac{|\mathcal{I}_{\omega, \eta, n}^c|}{(2(n-k)+1)^d}. \quad (4.4.2)$$

We now look for an upper bound for  $|\mathcal{I}_{\omega, \eta, n}^c|$ . If  $(T_x \omega)_{\Lambda_k} = p_k$  and  $(T_x \eta)_{\Lambda_k} \neq p_k$ , then  $\omega_y \neq \eta_y$  for at least one site  $y \in \Lambda_k + x$ . Such a  $y$  can produce as many as  $(2k+1)^d$  sites such that  $(T_x \omega)_{\Lambda_k} = p_k$  and  $(T_x \eta)_{\Lambda_k} \neq p_k$ . Hence

$$\begin{aligned} |\mathcal{I}_{\omega, \eta, n}^c| &\leq (2k+1)^d |\{x \in \Lambda_{n-k} : \omega_x \neq \eta_x\}| \\ &\leq (2k+1)^d |\{x \in \Lambda_n : \omega_x \neq \eta_x\}| \\ &\leq (2k+1)^d \bar{d}_{\Lambda_n}(\omega, \eta). \end{aligned}$$

Hence (4.4.2) yields

$$\|f_{n,k}(\omega, \cdot) - f_{n,k}(\eta, \cdot)\|_{TV} \leq \frac{(2k+1)^d}{(2(n-k)+1)^d} \bar{d}_{\Lambda_n}(\omega, \eta). \quad (4.4.3)$$

Obviously there exists  $\tilde{N} > k$  such that for all  $n \geq \tilde{N}$  we have

$$\left( \frac{2k+1}{2(n-k)+1} \right)^d \leq \frac{5}{4}$$

therefore, if we take

$$\bar{d}_{\Lambda_n}(\omega, \eta) \leq \frac{2\varepsilon}{5(2k+1)^d} (2(n-k)+1)^d$$

we finally obtain

$$\|f_{n,k}(\omega, \cdot) - f_{n,k}(\eta, \cdot)\|_{TV} \leq \frac{\varepsilon}{2}$$

for all  $n \geq \tilde{N}$ , which concludes the proof of the lemma.

### 4.4.3 A bound on relative entropy

Recall that “log” stands for the natural logarithm.

**Lemma 4.4.1.**

Let  $\nu$  and  $\mu$  be probability measures on a finite set  $A$ . Then

$$\sum_{a \in A} \nu(\{a\}) \left| \log \frac{\nu(\{a\})}{\mu(\{a\})} \right| \leq H(\nu|\mu) + \frac{2}{e} \quad (4.4.4)$$

where

$$H(\nu|\mu) = \sum_{a \in A} \nu(\{a\}) \log \frac{\nu(\{a\})}{\mu(\{a\})}.$$

*Proof.* Define

$$A^- = \left\{ a \in A : \log \frac{\nu(\{a\})}{\mu(\{a\})} < 0 \right\}.$$

Now

$$\begin{aligned} \sum_{a \in A} \nu(\{a\}) \left| \log \frac{\nu(\{a\})}{\mu(\{a\})} \right| &= \sum_{a \in A \setminus A^-} \nu(\{a\}) \log \frac{\nu(\{a\})}{\mu(\{a\})} + \sum_{a \in A^-} \nu(\{a\}) \log \frac{\mu(\{a\})}{\nu(\{a\})} \\ &= H(\nu|\mu) + 2 \sum_{a \in A^-} \nu(\{a\}) \log \frac{\mu(\{a\})}{\nu(\{a\})}. \end{aligned}$$

By the concavity of the logarithm function and Jensen’s inequality we get

$$\begin{aligned} \sum_{a \in A^-} \nu(\{a\}) \log \frac{\mu(\{a\})}{\nu(\{a\})} &= \nu(A^-) \sum_{a \in A^-} \frac{\nu(\{a\})}{\nu(A^-)} \log \frac{\mu(\{a\})}{\nu(\{a\})} \\ &\leq \nu(A^-) \log \frac{\mu(A^-)}{\nu(A^-)} \\ &\leq \nu(A^-) \log \frac{1}{\nu(A^-)} \leq e^{-1} \end{aligned}$$

where we used the elementary inequality  $-x \log x \leq e^{-1}$ ,  $x \geq 0$ . Therefore we arrive at (4.4.4).  $\square$

## 4.5 A final remark

We proved that (4.2.4) is a direct consequence of the Gaussian concentration bound, so we don’t need to use the blowing-up property. The drawback is that this proof uses the assumption  $|S| < \infty$ , whereas the one based on blowing-up also works when  $S$  is countably infinite. Take  $n$  and  $k$  such that  $n > k \geq 0$ . Consider the local function

$$F_{n,k}(\omega) = \|\mathfrak{f}_{n,k}(\omega; \cdot) - \mu_k\|_{TV}.$$

We have  $\delta_z(K_{n,k}) = 0$  whenever  $z \in \mathbb{Z}^d \setminus \Lambda_{n-k}$ . We now estimate  $\delta_x(K_{n,k})$ . Consider two configurations  $\omega$  and  $\eta$  in  $\Omega$  such that there exists  $x \in \Lambda_{n-k}$  such that  $\omega_y = \eta_y$  for all  $y \neq x$ , and  $\omega_x \neq \eta_x$ . It is easy to verify that

$$|F_{n,k}(\omega) - F_{n,k}(\eta)| \leq \|\mathfrak{f}_{n,k}(\omega; \cdot) - \mathfrak{f}_{n,k}(\eta; \cdot)\|_{TV}.$$



Using (4.4.3) we obtain (since  $\bar{d}_\Lambda(\omega, \eta) = 1$ )

$$\delta_x(K_{n,k}) \leq \frac{(2k+1)^d}{(2(n-k)+1)^d}. \quad (4.5.1)$$

We plug this bound to (3.2.3) to obtain

$$\mu\{\omega \in \Omega : F_{n,k}(\omega) \geq \mathbb{E}_\mu[K_{n,k}] + u\} \leq \exp\left(-\frac{(2(n-k)+1)^d u^2}{4D(2k+1)^d}\right) \quad (4.5.2)$$

for all  $u > 0$ . We now seek for an upper bound for  $\mathbb{E}_\mu[K_{n,k}]$ . We start by an upper bound for  $\mathbb{E}_\mu[|f_{n,k}(\cdot; p_k) - \mu([p_k])|]$ . For all  $\lambda > 0$ , we have

$$\mathbb{E}_\mu \left[ e^{\lambda |f_{n,k}(\cdot; p_k) - \mu([p_k])|} \right] \leq \mathbb{E}_\mu \left[ e^{\lambda (f_{n,k}(\cdot; p_k) - \mu([p_k]))} \right] + \mathbb{E}_\mu \left[ e^{-\lambda (f_{n,k}(\cdot; p_k) - \mu([p_k]))} \right]$$

where we simply decomposed according to the sign of  $f_{n,k}(\cdot; p_k) - \mu([p_k])$  using the indicator functions which are bounded above by one. Now we can apply GCB (since  $\mathbb{E}_\mu[f_{n,k}(\cdot; p_k)] = \mu([p_k])$ ) to get

$$\mathbb{E}_\mu \left[ e^{\lambda |f_{n,k}(\cdot; p_k) - \mu([p_k])|} \right] \leq 2e^{D\lambda^2 \ell_{n,k}^2}$$

where  $\ell_{n,k}^2 := 1/(2(n-k)+1)^d$ . We use the Jensen inequality and then divide out by  $\lambda > 0$  to get

$$\mathbb{E}_\mu[|f_{n,k}(\cdot; p_k) - \mu([p_k])|] \leq \frac{\log 2}{\lambda} + D\lambda \ell_{n,k}^2.$$

After optimizing over  $\lambda > 0$  we obtain

$$\mathbb{E}_\mu[|f_{n,k}(\cdot; p_k) - \mu([p_k])|] \leq 2\sqrt{D \log 2} \ell_{n,k}.$$

Therefore, we have

$$\mathbb{E}_\mu[K_{n,k}] \leq \sqrt{D \log 2} |S|^{(2k+1)^d} \ell_{n,k}.$$

After some simple algebra, we obtain from (4.5.2) that, given  $k$  and  $\varepsilon > 0$

$$\mu\{\omega \in \Omega : F_{n,k}(\omega) \geq \varepsilon\} \leq \exp\left(-\frac{(2n+1)^d}{4D2^d(2k+1)^d} \left(\varepsilon - \frac{\sqrt{D \log 2} |S|^{(2k+1)^d}}{(2(n-k)+1)^{\frac{d}{2}}}\right)^2\right)$$

provided that  $n$  is large enough, which ensures in particular that the difference inside the square is positive. Therefore, there exists  $\delta = \delta(\varepsilon, k)$  and  $N = N(\varepsilon, k)$  such that

$$\mu\{\omega \in \Omega : F_{n,k}(\omega) \geq \varepsilon\} \leq \exp\left(-(2n+1)^d \delta\right).$$

Then

$$\mathbb{E}_\mu[K_{n,k}] \leq \sqrt{D \log 2} |S|^{(2k+1)^d} \ell_{n,k}$$

for all  $n \geq N$ , which is exactly (4.2.4).

## Chapter 5

# Numerical study of concentration inequalities for the 2D Ising model

Our goal is to explore numerically concentration inequalities for a large class of observables. We will introduce the classical Metropolis algorithm to generate configurations according to the equilibrium distribution and thereafter to estimate the concentration constants involved at different temperatures. The first case was to consider the uniqueness regime in which a Gaussian concentration holds in the high-temperature regime with a concentration constant depending on the Dobrushin contraction coefficient (see [14]). Dropping down the temperature to the critical one, the model still exhibits a unique phase which satisfies strong mixing properties (see [28, 29, 61, 81, 80]) and, for this reason we aim to illustrate the intuition that Gaussian concentration bound holds for all  $\beta < \beta_c$ . We know, by contraposition, that at  $\beta = \beta_c$  the Ising model cannot satisfy a Gaussian concentration bound because it does not even satisfy a variance inequality (see 3.3). In the low-temperature regime, one cannot expect such a Gaussian concentration bound and the authors ([13]) proved in a certain low-temperature regime that a stretched-exponential concentration bound holds (see 3.2.6). Our result confirms this concentration type by stating that non-uniqueness prevents GCB to hold ([19]). Thanks to our simulations, we could depict the concentration constant behavior according to the temperature and envisage a possible extension of theoretical results. Indeed, for all temperatures above  $T_c$ , the concentration constant acts as a continuous and decreasing function of the temperatures diverging as a power-law at  $T_c$ . Consequently, a Gaussian concentration bound seems to hold for all temperatures above the critical one. When the temperatures drop down, we estimated the exponent quantifying the concentration decay according to the deviation by simulating the size behavior of the probability of a fluctuation of different observable. This result echoes the large deviation result proved by D. Ioffe in [43] in which he describes the correct asymptotic behavior of the magnetization.

### 5.1 Setting

Here and below, we will use the same setting. The configuration space is  $\Omega = S^{\mathbb{Z}^2}$ , where  $S = \{-1; +1\}$ . We define the distance between a subset

$\Lambda \subset \mathbb{Z}^d$  and  $x \in \mathbb{Z}^d$  as

$$d(\Lambda, x) = \inf\{\|x - y\|_\infty; y \in \Lambda\}$$

and we denote by  $\partial\Lambda$  the 1-boundary defined as follows

$$\partial\Lambda := \{x \in \Lambda^c; \|x - y\|_\infty \leq 1 \text{ with } y \in \Lambda\}. \quad (5.1.1)$$

Let  $\omega^{(x)} \in \Omega$  be the configuration obtained from  $\omega \in \Omega$  by flipping the spin at site  $x \in \mathbb{Z}^d$ , that is the configuration such that  $(\omega^{(x)})_y = \omega_y$  if  $y \neq x$  and  $(\omega^{(x)})_y = -\omega_y$  if  $y = x$ . Recall that we defined centered cubes this way: for every  $n \in \mathbb{Z}_+$ , let

$$\Lambda_n = \{x \in \mathbb{Z}^2 : -n \leq x_i \leq n, i = 1, 2, \dots, d\}.$$

Finally, the shift action  $(T_x, x \in \mathbb{Z}^2)$  is defined as follows: for each  $x \in \mathbb{Z}^d$ ,  $T_x : \Omega \rightarrow \Omega$  and  $(T_x \omega)_y = \omega_{y-x}$ , for all  $y \in \mathbb{Z}^2$ . We also recall what we mean by Ising model.

**Definition 5.1.1** (The 2D Ising model).

*The Ising model refers to the nearest-neighbor ferromagnetic Ising model in dimension two with positive external field in which the potential is defined as follows*

$$\Phi(\Lambda, \omega) = \begin{cases} -h\omega_x & \text{if } \Lambda = \{x\} \\ -J\omega_x\omega_y & \text{if } \Lambda = \{x, y\} \text{ and } \|x - y\|_1 = 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.1.2)$$

Later, we will rather consider the potential  $\beta\Phi$  where  $\beta \in \mathbb{R}_+$  is the inverse temperature. (Since the Ising model potential  $\Phi$  will be fixed, we will always refer to  $\beta$  for the potential  $\beta\Phi$ .) Therefore, the total energy inside  $\Lambda$  is given by the Hamiltonian

$$\mathcal{H}_\Lambda(\omega|\eta) = - \sum_{\substack{i, j \in \Lambda \\ \|i-j\|_1=1}} \omega_i\omega_j - \sum_{\substack{i \in \Lambda, j \in \partial\Lambda \\ \|i-j\|_1=1}} \omega_i\eta_j - h \sum_{i \in \Lambda} \omega_i.$$

we will denote by  $\gamma^\beta$  its corresponding specification (see 2.3.3 for the definition). We recall that for  $h = 0$  there exists an inverse temperature  $\beta_c$  such that for all  $\beta \leq \beta_c$ ,  $\mathcal{G}(\beta\Phi)$  is a singleton, and for all  $\beta > \beta_c$  there exists two different extreme Gibbs measures  $\mu_\beta^+$  and  $\mu_\beta^-$  obtained with the “+” and “-” boundary conditions. For  $h > 0$ , there exists a unique Gibbs measure at all temperature. More generally, let  $\Lambda \Subset \mathbb{Z}^d$ , we denote by  $\mu_\beta^\eta$  the Gibbs measure obtained with the boundary condition  $\eta \in \Omega_{\Lambda^c}$ . In this chapter, we will assume that  $h = 0$ .

## 5.2 Computations and estimates

### 5.2.1 Introduction

In this section, we will introduce an important Monte Carlo sampling method to evaluate the behavior of observables at every finite temperature. We will

give a brief glimpse of the relevance and effectiveness of this stochastic approach and show up numerical limitations we can encounter. Particularly for this model, the energy, spontaneous magnetization, magnetic susceptibility etc can be calculated exactly giving its critical exponent. Mostly, we don't know the analytic expression at finite temperature but we can obtain the critical exponent close to phase transition. For more details, we refer the reader to [51], [42].

## 5.2.2 Computation problems

Roughly speaking, the phase transition phenomena corresponds to an abrupt change in the behavior of observables by moving one parameter. Few of them are of comparable interest in the ferromagnetic Ising model in two dimensions as magnetization and local energy because they are easy to compute and we know analytically their expression. In fact, given a boundary condition of the system, we can calculate the probability for the system to be in this specified configuration and we can also derive easily the macroscopic quantities. So, we are interested in the expected value of magnetization, the local energy, the magnetic susceptibility and heat capacity in a finite box  $\Lambda$ .

### Definition 5.2.1.

Let  $\Lambda \in \mathbb{Z}^2$ ,  $\omega \in \Omega$  and  $\eta \in \Omega$  the boundary condition. We denote respectively the magnetization, the average magnetization, the local energy, the average local energy, the magnetic susceptibility and the heat capacity in  $\Lambda$  by

$$m_{\Lambda}^{\eta}(\omega) = \sum_{i \in \Lambda} \omega_i, \quad \mathbb{E}_{\mu_{\beta}^{\eta}}(m_{\Lambda}^{\eta}) = \sum_{\omega \in \Omega} m_{\Lambda}^{\eta}(\omega) \gamma_{\Lambda}^{\beta}(\omega | \eta),$$

$$E_{\Lambda}^{\eta}(\omega) = \frac{1}{2} \sum_{\substack{i \in \Lambda, j \in \Lambda \cup \partial \Lambda \\ \|i-j\|=1}} \omega_i \omega_j, \quad \mathbb{E}_{\mu_{\beta}^{\eta}}(E_{\Lambda}^{\eta}) = \sum_{\omega \in \Omega} E_{\Lambda}^{\eta}(\omega) \gamma_{\Lambda}^{\beta}(\omega | \eta),$$

$$\chi_{\Lambda}^{\eta} = \beta \left[ \mathbb{E}_{\mu_{\beta}^{\eta}}((m_{\Lambda}^{\eta})^2) - \mathbb{E}_{\mu_{\beta}^{\eta}}(m_{\Lambda}^{\eta})^2 \right], \quad \mathcal{C}_{\Lambda}^{\eta} = \beta^2 \left[ \mathbb{E}_{\mu_{\beta}^{\eta}}((E_{\Lambda}^{\eta})^2) - \mathbb{E}_{\mu_{\beta}^{\eta}}(E_{\Lambda}^{\eta})^2 \right].$$

For  $\omega \in \Omega$ ,  $x \in \mathbb{Z}^2$  and  $\Lambda \in \mathbb{Z}^2$  we also introduce respectively, correlations  $\mathbb{E}_{\mu_{\beta}^{\eta}}(\omega_0 \omega_x)$  and the empirical pair correlations

$$\Gamma_{\Lambda, x}(\omega) = \frac{1}{|\Lambda|} \sum_{y \in \Lambda} \omega_y \omega_{x+y}.$$

Further, if it's sufficiently clear, we omit the size and boundary condition indexes in observables.

**Remark 5.1.** In this model, there exists an analytical expression of the partition function from which every observable we consider follows by derivation (see [67]).

There are different ways to simulate the system at equilibrium and calculate the average of an observable. The intuitive one is to calculate the magnetization and the energy for each configuration and weighting them by the probability factor to calculate the expected value. But, this method has important limiting aspects, considering boxes of very large size (thermodynamic

limit) the calculus reveals to be impossible due to the  $2^{|\Lambda|}$  possible different configurations. A good way to do these calculations is to choose configurations with the same probability factor and give them the same weight. This is known as the Metropolis algorithm. However, in the low-temperature regime, this algorithm tends to freeze the system in only few configurations and estimating particular events becomes impossible. We tried without significant improvement to get better access to these computational estimations by a Modified Metropolis algorithm based on subset simulation techniques (see [85, 5]).

### 5.2.3 Classical Metropolis algorithm

This algorithm works by creating a sequence of sample configurations in such a way that, as more and more samples are produced, the distribution of configuration more closely approximates the desired equilibrium state. More precisely, We generate a sample configuration from a previous one using a transition probability which depends on the energy change between the initial configuration and the final one. In this chapter, we denote by  $\Lambda \in \mathbb{Z}^2$  the centered “square” of side length  $L \in 2\mathbb{Z}_+$  such that  $|\Lambda| = L^2$ . We also denote by  $P_\Lambda^\eta(\omega, \omega')$  the transition probability from the configuration  $\omega \in \Omega$  to  $\omega' \in \Omega$  with boundary condition  $\eta \in \Omega$  such that  $\omega_{\Lambda^c} = \omega'_{\Lambda^c} = \eta_{\Lambda^c}$ . We want to be sure that the final state corresponds to the equilibrium state, so we impose to the probability measure to satisfy the following detailed balance equation

$$\gamma_\Lambda^\beta(\omega|\eta)P_\Lambda^\eta(\omega, \omega') = \gamma_\Lambda^\beta(\omega'|\eta)P_\Lambda^\eta(\omega', \omega).$$

Thus, going from the configuration  $\omega$  to  $\omega'$  and vice-versa depends only on the energy difference between these two configurations  $\Delta\mathcal{H}_\Lambda = \mathcal{H}_\Lambda(\omega'|\eta) - \mathcal{H}_\Lambda(\omega|\eta)$ .

$$\frac{P_\Lambda^\eta(\omega, \omega')}{P_\Lambda^\eta(\omega', \omega)} = \exp(-\beta\Delta\mathcal{H}_\Lambda).$$

At this time, we need to specify uniquely the transition probability from one configuration to another. To do that we use the Metropolis form of the transition probability:

$$P_\Lambda^\eta(\omega, \omega') = \begin{cases} \exp(-\beta\Delta\mathcal{H}_\Lambda) & \text{if } \Delta\mathcal{H}_\Lambda < 0 \\ 1 & \text{otherwise} \end{cases}.$$

Hence, starting from an initial distribution of configurations and applying the transition probability  $P_\Lambda^\eta$  generates a Markov process in which the initial distribution converges to the equilibrium state when the time step diverges to infinity. In a computational way, we have the following algorithm

1. We choose randomly a spin from an initial configuration.
2. We flip this spin and calculate the energy difference between this configuration and the previous one.
3. If the difference is negative, keep this spin flipped and restart the process. If not go to the next step.
4. Generate a random number  $r \in (0, 1)$ . If  $\exp(-\beta\Delta\mathcal{H}_\Lambda) > r$ , then keep the spin flipped. If not let it in its initial state.

5. return.

We call one step of this algorithm a micro Monte Carlo step (mMCs) and the standard measure of one unit Monte Carlo time  $t_{MC}$  is defined by the mean number of mMCs needed to change every spins also called MCs. At fixed temperature, before collecting data, we need to perform a large number of iterations of the micro MC step in order to generate configurations with a probability measure converging to the unique equilibrium state. The choice of a reasonable amount of mMCs and the precision of the results depend upon many factors as temperature, size of the lattice, boundary conditions etc and will be discussed in the following sections.

#### 5.2.4 Finite-size effect

Let's recall that our main interest is to determine statistical properties of the infinite-volume Gibbs measures. Since we can't numerically consider infinite lattices, we can rightly ask how finite-size effects act on the precision of the measurement on observables. To observe these effects, one may compute a large number of configurations distributed according to finite-volume Gibbs measures with different size length and simulate the different behavior of an observable with respect to the volumes (see [49]). By this method, one can approximate and extract numerically the critical behavior of observables according to the size of the system (or equivalently to the temperature). For example, for very small lattices, one can miss the essential phenomena which is a drastic change in the behavior of the system at critical temperature. For this reason we need to choose a sufficiently large lattice. Another source of errors might emerge from the implementation of large systems and is related with the correlation between two successive configurations.

#### 5.2.5 Finite sampling time effect

Due to physical limits of computing, one has to choose a convenient way of using the computer resources by simulating systems with different size during different time scales. Here, we give some solutions to reduce statistical errors produced by small number of MCs considering large size lattices. We need to consider times large enough to produce typical configurations from the equilibrium state. We refer the reader to D. Landau and K. Binder in [51] where they methods to study the transient "phase" evolution starting from a random configuration and quantify the time we need to simulate these configurations. Roughly speaking, they call relaxation time  $\tau_e$ , the time when the transient phase ends. The main computational problem that we have to deal with is that the influence of correlation between configurations intensifies close to the critical temperature and the relaxation time behaves as:

$$\tau_e \sim L^{z_e}$$

where  $z_e$  is a real number smaller than 2. Beyond transient phase, other typical configurations and averages are produced by computing the Metropolis algorithm but they always exhibit temporal correlations responsible for errors if the number of MCs is too small. Since two configurations are highly correlated applying a micro Monte Carlo step, the author in [51] could quantify the behavior of an autocorrelation function which indicates, for large time,

how long we have to “wait” a decorrelation of samples to ensure its statistical quality. Typically, far from the critical temperature, the characteristic correlation time  $\tau_c$  depending on the temperature is attained in a few MCs but at the critical point it diverges with the size of the lattice:

$$\tau_c \sim L^z$$

where  $z \approx 2, 2$  and in this model,  $z_e$  is always smaller than  $z$  depending on the observable. Particularly, close to the critical temperature, the number of MCs must be larger than the square size of the lattice to decorrelate two statistical measures. This means also, that far from this point it's not necessary to compute with large times because it will produce the same results.

### 5.3 Estimation of concentration constants

We have arrived now at the core of this chapter: the estimation of concentration constants. Once the model has been simulated, we collect, from the typical configurations, the fluctuations of different observables around their mean to produce an histogram that we will fit with the maximum likelihood density function. At the end, we compare our estimations about magnetization, energy and concentration constants to the analytical results already calculated.

#### 5.3.1 Method

We need to introduce some notation to quantify numerically and clarify the difference between the observable we measure by simulating the model and their theoretical expectation, the expectation and the fluctuations of a local function  $K : \Omega \rightarrow \mathbb{R}$ . Let  $i \in \mathbb{N}$ ,  $\omega^{(i)} \in \Omega$  the  $i$ -th sample configuration and  $K(\omega^{(i)})$  the  $i$ -th sample of the simulated observable  $K$  (in our context, the exponent  $i$  can also acts as a variable of time in term of MCs between each sample). We denote by

$$\langle K \rangle_s = \frac{1}{s} \sum_{i=1}^s K(\omega^{(i)})$$

the empirical mean of the observable  $K$  which, by the ergodic theorem, converges to  $\mathbb{E}_{\mu_\beta}(K)$ . Then one may compute a large number of samples to obtain a precise approximation of fluctuations. Typically, we computed  $s = 10^5$  to get closer to the probability distribution in a convenient amount of time. Once done, with R-studio, a classical fitting procedure provides by successive tests, the maximum likelihood function coinciding with the empirical measure of fluctuation. In fact, it approximates the empirical probability distribution behavior of fluctuations by the expected real-density function and we deduce the simulated concentration constant  $D_{\beta, \Lambda, s}^{(emp)}(K)$  by its classical parameters. With this notation we define the empirical susceptibility and the empirical heat

capacity and their respective estimators by

$$\begin{aligned}\chi_{\beta,\Lambda,s}^{(emp)} &= \beta \left( \langle m_\Lambda^2 \rangle_s - \langle m_\Lambda \rangle_s^2 \right) \\ \chi_{\beta,\Lambda,s}^{(est)} &= \beta (m_\Lambda(\omega))^2 - \langle m_\Lambda^2 \rangle \\ \mathcal{C}_{\beta,\Lambda,s}^{(emp)} &= \beta^2 \left( \langle E_\Lambda^2 \rangle_s - \langle E_\Lambda \rangle_s^2 \right) \\ \mathcal{C}_{\beta,\Lambda,s}^{(est)} &= \beta^2 \left( (E_\Lambda(\omega))^2 - \langle E_\Lambda^2 \rangle \right).\end{aligned}$$

### 5.3.2 Settings and preliminary results

The program implemented gives the behavior of observables in a “small” box  $\Lambda$  of different side length  $L = 100$  contained in a box of side length  $L = 500$  at different range of temperatures from  $T = 1.50$  to  $T = 5.0$  with many points concentrated around  $T_c \approx 2,269$ . Far from the critical temperature, we produce configurations at equilibrium computing generously  $10^5$  Monte Carlo steps and to decorrelate each sample we compute 10 MCs. Very close to the critical temperature ( $[2.255, 2.275]$ ), the correlation in time and in space increases drastically and we cannot reach a very accurate result in a reasonable amount of time. Nevertheless, we successfully approach theoretical results even in this short range of temperature. To support this statement, we plotted several classical observables below (see figures) and their expected behavior.

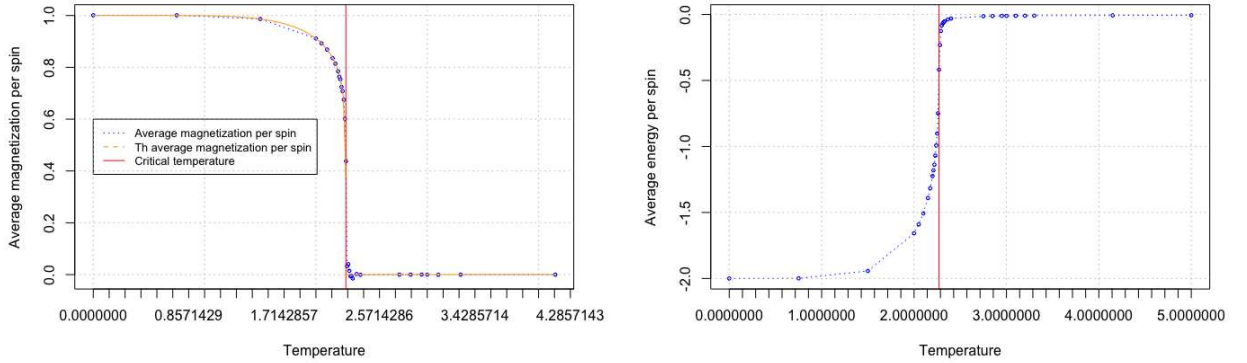


Figure 5.1: This plot shows the result of Average Magnetization and Average Energy per spin.

From the study of J. Kotze (see [49]), we observe the emergence of a critical behavior close to  $T_c$  for the average magnetization and the average energy per spin curves for different box sizes whenever the size increases. In fact, for small size lattices, the curve is smooth but for  $L = 100$ , their behavior coincides almost exactly with the theoretical expression at all temperatures (see 5.1) :

$$\mathbb{E}_{\mu_\beta}(m^+) = \begin{cases} \left[ 1 - \frac{1}{\sinh^4(2\beta)} \right]^{\frac{1}{8}} & \text{if } \beta > \beta_c \\ 0 & \text{otherwise} \end{cases}.$$



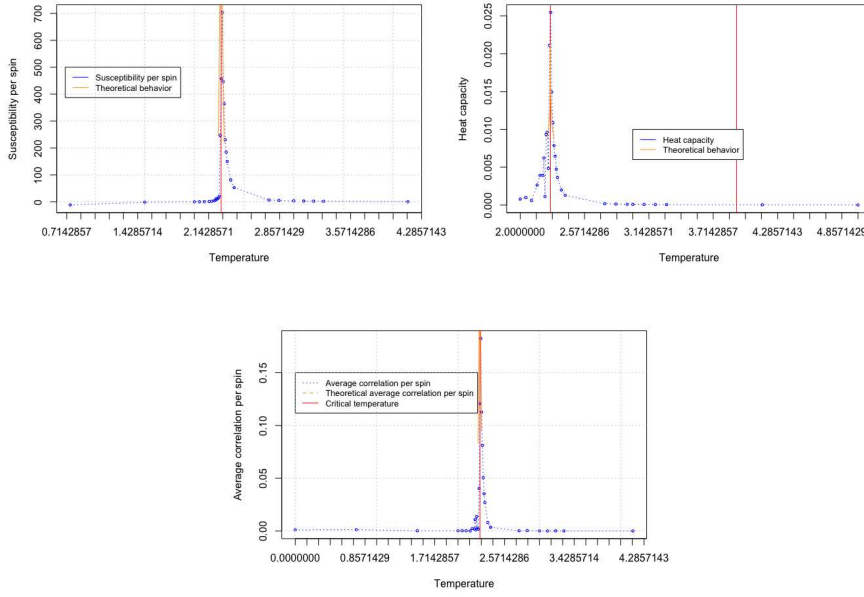


Figure 5.2: These plots show the result of Susceptibility per spin, Specific Heat per spin and correlation at different temperature.

The susceptibility, heat capacity and correlations also emphasize this phenomenon. For these three curves, we notice that the divergence would occur at a phase transition but, with finite-size lattices, there emerges a peak value at the hypothetic divergence point. Even if this limitation leads to some errors, the critical behavior follows, as expected, the power-law divergence with the right critical exponent (see 5.2). The following sections deal with the estimation of such quantities and the behavior of concentration constants at all finite temperature.

### 5.3.3 Gaussian concentration constant estimation

Let's recall that a Gaussian concentration property holds under a sufficiently high-temperature condition (see 3.2.1).

**Theorem 5.3.1** ([13, 50]).

For all  $\beta < \underline{\beta} = \frac{1}{2} \ln \frac{5}{3}$ , the unique Gibbs measure  $\mu_\beta$  satisfies GCB( $D$ ) with  $D = 2(1 - \tanh(\beta))^{-2}$ .

In this section, we fix the side length of the box  $L = 100$ . At high temperature, one expects a Gaussian concentration bound for every sufficiently regular observable  $K : \Omega \rightarrow \mathbb{R}$ . For this reason, we fit numerically the histogram of fluctuations at different temperature until the critical one (see figure 5.3) with a Gaussian density function. More precisely, we fit

$$\frac{1}{s} \sum_{i=1}^s \mathbb{1}_{\{\omega \in \Omega: K(\omega^{(i)})(\omega) - \langle K \rangle_s \geq u\}}$$

by the maximum likelihood function, namely

$$\exp\left(-\frac{(u - \langle K \rangle_s)^2}{2(\langle K^2 \rangle_s - \langle K \rangle_s^2)}\right).$$

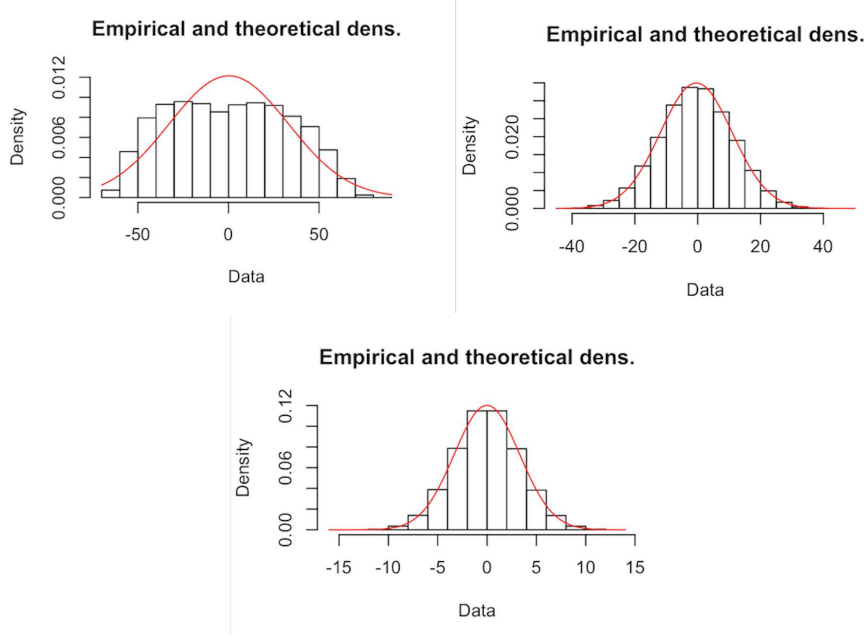


Figure 5.3: These graphs show qualitatively the fitting procedure of the fluctuations of the magnetization at different temperature ( $T = 2.29, 2.40, 3.00$ ).

By simple relations, we derive the expression of  $D_{\beta, \Lambda, s}^{(emp)}(K)$  from the variance  $\langle K^2 \rangle_s - \langle K \rangle_s^2$  and its behavior for different observables at finite temperatures (see 5.3) as follows

$$D_{\beta, \Lambda, s}^{(emp)}(K) = \frac{\langle K^2 \rangle_s - \langle K \rangle_s^2}{2\|\underline{\delta}(K)\|_2^2}.$$

**Remark 5.2.** For our estimations, we will need the following values of  $\|\underline{\delta}(K)\|_2^2$ .

$$\begin{aligned} \|\underline{\delta}(m)\|_2^2 &= 4|\Lambda|, & \|\underline{\delta}(m^2)\|_2^2 &= 16(|\Lambda| - 1)^2|\Lambda|, \\ \|\underline{\delta}(E)\|_2^2 &= 16|\Lambda|, & \|\underline{\delta}(E^2)\|_2^2 &= 16(|\Lambda| - 1)^2|\Lambda|, \\ \|\underline{\delta}(\Gamma)\|_2^2 &= \frac{16}{|\Lambda|}. \end{aligned}$$

In figure 5.4, we computed the concentration constants of classical observables in a finite box of length  $L = 100$ . We represented by a vertical red lines, the critical temperature  $T_c$  and by a vertical yellow line the temperature  $\bar{T}$  above which Theorem 3.2.1 holds. We also computed the theoretical (when  $\bar{T} < T$ ) and hypothetical (when  $T < \bar{T}$ ) behavior of the Gaussian concentration constant and observe that  $\text{GCB}(2(1 - \tanh(\beta))^{-2})$  holds for a larger range of temperature with an optimal constant. In addition to illustrating that  $\text{GCB}(D_\beta)$  holds until the critical temperature, we also depict the divergence of concentration constants of interest. In fact, by this calculation method, we expect that the characteristic divergence as a power-law at critical temperature extends to all concentration constants as it does for the susceptibility

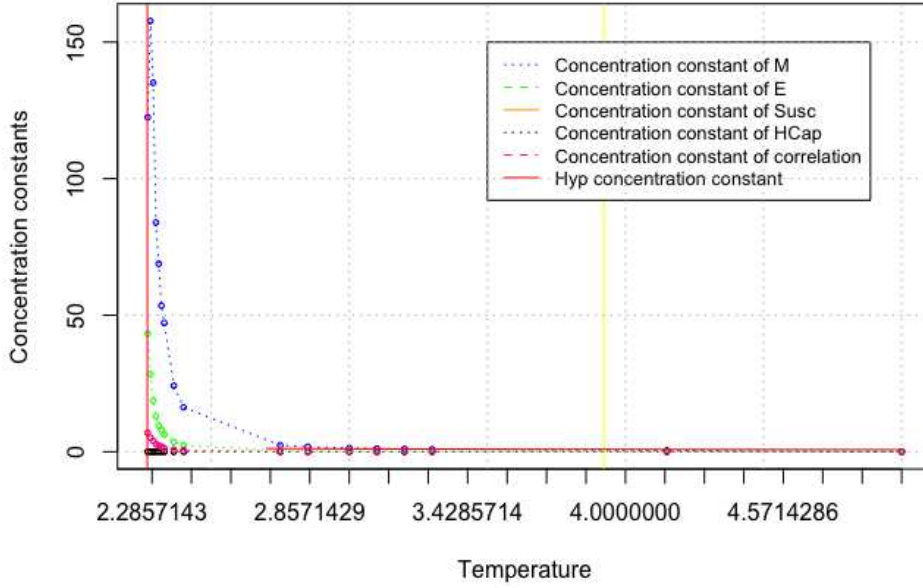


Figure 5.4: This plot shows the differing results of concentration constants depending on each observables and the hypothetical one.

and the concentration constant of the magnetization. In the next section, we analyze the power-law divergence of concentration constants by estimating their critical exponents.

### 5.3.4 Determination of concentration constant behavior

As we mentioned before in the finite-size effect section, it would be useful to introduce critical exponents to understand the nature of the divergence close to  $T_c$ . Let's introduce the critical exponent definition  $\zeta = \lim_{T \rightarrow T_c} \frac{K_{T-T_c}}{|T-T_c|}$  or by abuse of notation  $K_{T-T_c} \sim |T - T_c|^\zeta$ . To each concentration constant, we associate its critical exponent as follows:

$$D_{\beta,\Lambda,s}^{(emp)}(m_\Lambda) \sim |T - T_c|^{-a}, \quad D_{\beta,\Lambda,s}^{(emp)}(E_\Lambda) \sim |T - T_c|^{-b},$$

$$D_{\beta,\Lambda,s}^{(emp)}(m_\Lambda^2) \sim |T - T_c|^{-c}, \quad D_{\beta,\Lambda,s}^{(emp)}(E_\Lambda^2) \sim |T - T_c|^{-d}.$$

By extrapolating these behaviors with a large size length, the classical study of divergence consists in plotting the log-log graph which should provide a straight line with a coefficient slope equal to the critical exponent (see figures 5.5). To be sure of the reliability of this method, we gave precise estimations of known exponent values in the sense that the theoretical value belongs to the confidence interval of the estimation.

We collect all the calculation results we made in the following table.

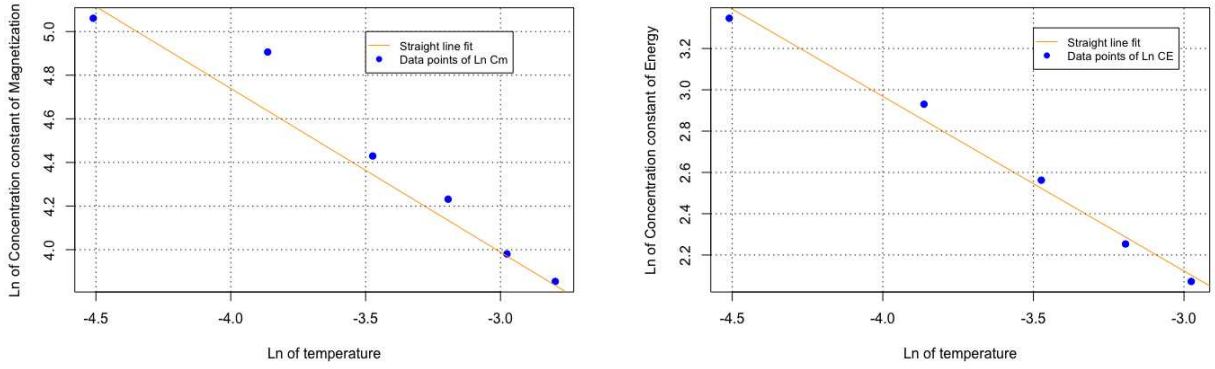


Figure 5.5: These plots show the linear fitting of the *log-log* concentration constant of the magnetization and energy with temperature.

Quantity	Exponent	Estimation	Theoretical value
Magnetization	$\delta$	$0.1260 \pm 0.0052$	1/8
Susceptibility	$\gamma$	$1.7877 \pm 0.1611$	7/4
Heat capacity	$\alpha$	0	/
Correlations	$\nu$	$1.0474 \pm 0.1247$	1
Dmagn	$a$	$0.7503 \pm 0.089$	3/4
Dener	$b$	$0.8456 \pm 0.0531$	?
Dsusc	$c$	$1.0195 \pm 0.1886$	?
Dcap	$d$	$2.0833 \pm 0.1745$	?
Dcorr	$f$	$0.69470 \pm 0.07052$	?

From the best of my knowledge, the theoretical values of the concentration constants are not known. For this reason, we denote by “?” the corresponding theoretical value. A special case in this table concerns Heat capacity in which the graph is not fitted by a straight line. The reason is because the exponent value  $\alpha$  is zero and its behavior has to be interpreted as logarithmic function of  $T - T_c$ . One may not be surprised by the susceptibility value and the concentration constant value of magnetization because there exists a direct relation of proportionality between them:

$$D_{\beta,\Lambda,s}^{(emp)}(m_\Lambda) = \frac{\chi_{\beta,\Lambda,s}^{(emp)}}{8\beta}.$$

To sum up, we provided numerically that a GCB( $D_{\beta,\Lambda,s}^{(emp)}$ ) holds for linear observables (magnetization) and non-linear ones (local energy, susceptibility, heat capacity) at all temperatures above  $T_c$ . Considering  $D_{\beta}^{(hyp)}$  as the supremum of all concentration constants, it may behave as the concentration constant of the magnetization.

### 5.3.5 Stretched-exponential concentration constant estimation

In the low-temperature regime, because of the existence of multiple Gibbs states, the Gaussian concentration cannot hold. Nevertheless, at sufficiently low temperature, one can control all moments 3.2.7 and deduce that the measure  $\mu_\beta^+$  satisfies a stretched-exponential concentration bound (see [13]). For the reader's convenience we recall Theorem 3.2.6.

**Theorem 5.3.2 ([13]).**

Let  $\mu_\beta^+$  be the “+”-phase of the low-temperature Ising model defined as above. There exists  $\bar{\beta} > \beta_c$  such that for all  $\beta > \bar{\beta}$  there exist  $\rho = \rho(\beta) \in (0, 1)$  and a constant  $D_\rho > 0$  such that, for all  $K \in \mathcal{L}$ , one has for all  $u > 0$

$$\mu_\beta^+ \left( \omega \in \Omega : |K(\omega) - \mathbb{E}_{\mu_\beta^+}(K)| \geq u \right) \leq 4 \exp \left( -\frac{u^\rho}{D_\rho \|\underline{\delta}(K)\|_2^\rho} \right).$$

The goal of this section is to estimate the parameters of the stretched-exponential concentration below the critical temperature. Thus, we follow optimistically the same path by making histograms of observables of interest. By comparison with classical density functions, the Weibull density function seems to have the maximum likelihood fit with the probability to observe a fluctuation of an observable. It's governed by two parameters: a shape parameter corresponding to the exponent in the concentration function and a scale parameter corresponding to the concentration constant. Unfortunately, fitting with such distributions generate exponents which do not agree with large-deviation results concerning the magnetization. In fact, the concentration bound, valid for all observables, might be overly pessimistic. So, instead of fixing the size of the box and calculating the deviation, we figure out the best exponent for every observable fixing the deviation parameter and compute the large-deviation function with respect to the size of the box. Formally, for fixed observables  $K : \Omega \rightarrow \mathbb{R}$  and deviations  $u \in \mathbb{R}$  we want to evaluate the large-deviation exponent  $q \in \mathbb{R}_+$  such that

$$\lim_{L \rightarrow \infty} \frac{1}{L^q} \log \mu_\beta^+ \left( K(\omega) - \mathbb{E}_{\mu_\beta^+}(K) \geq u \right) = -C(K, u) \quad (5.3.1)$$

where  $C(K, u) : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Thus, the determination of the optimal concentration exponent  $\rho$  relies totally on the large and moderate deviation analysis and its corresponding stretched-exponential expected behavior. Technically, we compute the number of occurrences of fluctuations of observables with respect to different sizes of the box and by an homogeneity argument of the concentration bound, we will guess its hypothetic value.

Fixing  $T = 1.50$  and  $T = 2.26$ , we plot the log-deviation probability of the magnetization, energy and correlations with respect to the the size length of the box up to a varying exponent. By a linear regression according to  $L^{0.1}$ ,  $L^{0.5}$ ,  $L$ ,  $L^{1.5}$  or  $L^2$  and by fixing the intercept coefficient to 0, we adjusted the slope of each curve and selected the one with the smallest standard error. We collect the coefficient of determination or the “Rsquare” value in tables and denote by “•” the highest value for the corresponding deviation and  $L^q$  with  $q = 0.1, 0.5, 1, 1.5, 2$ . Finally, with the same simulation techniques and for fixed deviations and the adjusted  $L^q$ , we evaluate the concentration constant  $D_\rho$  in this range of temperatures.

## Concentration exponent of the magnetization

For the plus-phase, Schonmann in [76] gave the correct asymptotic exponential decay to zero of the probability that the block spin magnetization deviates from its mean for sufficiently low temperature. This large deviation principle was refined by D. Ioffe who proved in [43] that a large and moderate deviation principle holds for small deviations. Let  $\omega \in \Omega$  and  $\Lambda \in \mathbb{Z}^2$ , we set  $M_\Lambda^+(\omega) = m_\Lambda^+(\omega)/|\Lambda|$  be the magnetization per spin. These results are as follows.

**Theorem 5.3.3** ([76],[43]).

Let  $\beta_c < \beta$  large enough. For  $u > 0$  we have

$$\lim_{L \rightarrow \infty} \frac{1}{L^2} \log \mu_\beta^+ \left( M_\Lambda^+(\omega) - \mathbb{E}_{\mu_\beta^+} \left( M_\Lambda^+ \right) \geq u \right) = -I(u) \quad (5.3.2)$$

where  $I : \mathbb{R} \rightarrow \mathbb{R}$  is the rate function defined as the Legendre-Fenchel transform of

$$p(h) := \lim_{L \rightarrow \infty} \frac{1}{L^2} \log \sum_{\omega \in S^\Lambda} \exp \left( \mathcal{H}_\Lambda(\omega) + h \sum_{\omega \in S^\Lambda} \omega_i \right).$$

For  $u \in \left[ -\mathbb{E}_{\mu_\beta^+} \left( M_\Lambda^+ \right), \mathbb{E}_{\mu_\beta^+} \left( M_\Lambda^+ \right) \right]$  we have

$$\lim_{L \rightarrow \infty} \frac{1}{L} \log \mu_\beta^+ \left( M_\Lambda^+(\omega) \leq u \right) = -C (\alpha(u))^{1/2} \quad (5.3.3)$$

with  $\alpha(u) = \left( \mathbb{E}_{\mu_\beta^+} \left( M_\Lambda^+ \right) - u \right) / \left( 2\mathbb{E}_{\mu_\beta^+} \left( M_\Lambda^+ \right) \right)$ .

$u$ \backslash T=1.50	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.009	0.9127405	0.9556466	0.9874373	• 0.9973909	0.9903635
-0.008	0.9190758	0.9611203	0.9912486	• 0.9990756	0.9896748
-0.007	0.9368466	0.9733068	0.9962118	• 0.9971406	0.9815893
-0.006	0.9486626	0.9808296	• 0.9983899	0.9944041	0.9746247
0.005	0.9329949	0.971329	0.9959632	• 0.9978508	0.9825328
0.006	0.9363763	0.9735209	0.996925	• 0.9979399	0.9821095
0.007	0.8985363	0.9462002	0.9835563	• 0.9981665	0.9946289

$u$ \backslash T=2.00	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.03	0.9499033	0.9818769	• 0.9990378	0.9945031	0.9740661
-0.02	0.9663745	0.991229	• 0.9995210	0.9868271	0.9593728
-0.01	0.9866349	• 0.9991084	0.9925001	0.9665669	0.9280966
-0.005	0.9973652	• 0.9982964	0.9778755	0.9397386	0.8911712
0.005	0.9919304	• 0.9996971	0.9874712	0.9565912	0.9140434
0.01	0.9731002	0.994557	• 0.9985993	0.9819703	0.9510996
0.02	0.9390434	0.9750328	0.9972354	• 0.9973867	0.9810323

$u$ \ T=2.24	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.1	0.997434	● 0.9975708	0.976049	0.9367774	0.8871148
-0.02	● 0.9987025	0.9941566	0.9675034	0.9241475	0.8714876
-0.01	● 0.9991225	0.9919115	0.9623082	0.9165803	0.8621642
-0.005	● 0.9993307	0.9896280	0.9574631	0.9098901	0.8543193
0.09	0.9616109	0.9885029	● 0.9987808	0.9873458	0.9604515
0.1	0.9542204	0.9844559	● 0.9989513	0.9914519	0.9679959

$u$ \ T=2.26	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.7	0.9407818	0.9640303	● 0.9723747	0.9618152	0.9381304
-0.6	0.9815318	● 0.9903771	0.9799396	0.9511663	0.9108743
-0.01	● 0.9984209	0.9855914	0.9493846	0.8979563	0.8389526
-0.005	● 0.9981291	0.984627	0.9476863	0.8956756	0.836245
0.1	0.9825165	● 0.9943309	0.9869233	0.9602785	0.9212302
0.15	0.960083	0.984179	● 0.9917334	0.9785711	0.950919
0.2	0.9445357	0.9753598	● 0.9909604	0.9848538	0.9629509

These tables describe the evolution of the order of the exponential decay 5.3.3 at low temperature in different regions of deviations. Despite the lack of information about extremely rare events, we could evaluate this probability for low temperatures  $T = 1.50$  until  $T = 2.26$ .

Close to  $T_c$ , for large deviations, the general behavior is a function of  $L^{1.0}$  while for small deviations it acts as a function of  $L^{0.1}$  and  $L^{0.5}$ . In fact, in this narrow range of deviation, we do not observe a clear difference between the  $L^{0.1}$  and  $L^{0.5}$ -dependence with respect to each other.

At  $T = 2.00$ , for small absolute values of  $u$ , the  $L^{0.5}$ -dependence dominates whereas for large  $u$  the simulation indicates a  $L^1$  and  $L^{1.5}$  dependence of the deviation probability.

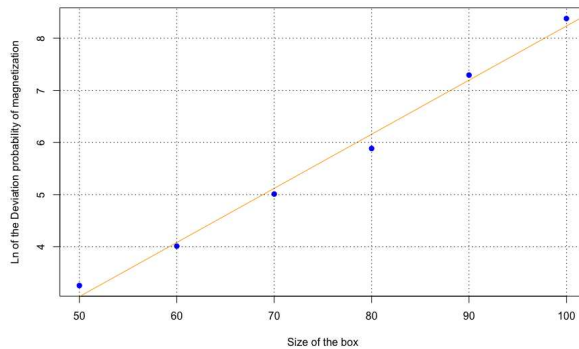


Figure 5.6: This plot shows the log-deviation probability of the magnetization at  $T = 2.00$  for  $u = 0.01$  with respect to  $L$ .

Dropping the temperature to  $T = 1.50$ , there is an apparent difference. First of all, we notice the expected decay of large fluctuations due to the difficulty of the algorithm to create diversified configurations. In this range of small deviations, we observe a surface-order large deviations which remind the theoretical large deviation result for the magnetization (see 5.3.3). For larger  $u$ , the  $L^{1.5}$ -dependence dominates. If we now were to think about

lower temperatures, regarding the complete evolution in these tables, the  $L^2$ -dependence increases more and more when the absolute value of  $u$  increases and the temperature goes down. Such an observation echoed the classical large deviation result for the magnetization (see 5.3.3). Combining the previous results for all low temperatures, it seems that the lowest exponent for the magnetization is  $q = 0.1$ . For sufficiently low temperatures, this value increases and corresponds to the expected large deviation results namely  $q = 0.5$ . It follows that, for sufficiently low temperature, the optimal value of the stretched-exponential concentration exponent might be valid for all deviations and is given by a relation with the lowest size-exponent of these tables namely  $\rho = 2q = 1$ .



## Concentration exponent of the energy

The large deviation principle 5.3.3 is only valid for functions having an additivity property like the magnetization. For general functions, such a theorem may be difficult to prove but we can illustrate a large deviation property for energy by computation without extra effort. Following the same path, we compute large and moderate deviation probabilities for the energy to determine the concentration exponent  $\rho$ .

$u$ \ T=1.50	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.035	0.9580257	0.986614	● 0.9995497	0.991098	0.9672438
-0.025	0.9804676	● 0.9976540	0.9963408	0.9747378	0.9395377
-0.01	0.9965219	● 0.9987139	0.9798727	0.9431922	0.8958744
-0.005	● 0.9991771	0.9957245	0.9700556	0.9272674	0.874827
-0.0025	● 0.9996904	0.9936813	0.9651976	0.9201781	0.8661303
0.025	0.9957684	● 0.9992405	0.9817242	0.9460219	0.8993549
0.05	0.9885047	● 0.9996988	0.9913282	0.963605	0.9234442
0.075	0.9837442	● 0.9985359	0.9944841	0.9706104	0.9336852

$u$ \ T=2.00	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.15	0.9653277	0.9905494	● 0.9993034	0.987054	0.9600147
-0.1	0.980303	● 0.9976033	0.9964384	0.9749778	0.9399052
-0.01	● 0.9999690	0.9924332	0.9623015	0.9159827	0.86098
-0.005	● 0.9999635	0.9905023	0.9582055	0.9101062	0.8537426
-0.0025	● 0.9998490	0.9892065	0.9555697	0.9063559	0.8491291
0.2	0.9865098	● 0.9993602	0.9929177	0.9668479	0.9280141
0.25	0.9850756	● 0.9990349	0.9939231	0.9690299	0.9311679
0.3	0.9831376	● 0.9982978	0.9946785	0.9711792	0.9345584

$u$ \ T=2.24	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.45	0.961725	0.9886012	● 0.9993348	0.9887889	0.9630712
-0.4	0.9670276	0.9915858	● 0.9994338	0.9862419	0.9582635
-0.01	● 0.9999011	0.9914629	0.9604081	0.9134132	0.8579591
-0.005	● 0.9998932	0.9906908	0.9587665	0.9110468	0.8550308
-0.0025	● 0.9998822	0.9904110	0.9581804	0.9102054	0.8539919
0.35	0.9922145	● 0.9997601	0.9871122	0.9556938	0.9125437
0.4	0.9912382	● 0.9997059	0.9881811	0.957769	0.9154593
0.45	0.9898673	● 0.9994782	0.9893621	0.960233	0.9190164

$u$ \ T=2.26	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.6	0.9506218	0.9814776	● 0.9966774	0.9897603	0.9667559
-0.5	0.9568687	0.9849517	● 0.9967782	0.9868322	0.9612827
-0.005	● 0.9994303	0.9927852	0.9637359	0.9183629	0.8641096
-0.0025	● 0.9994673	0.992608	0.9633178	0.9177465	0.8633411
1	0.9481525	0.9776994	● 0.9919403	0.9847581	0.9620304
1.1	0.9395303	0.9721972	● 0.9901206	0.9861035	0.9658824
1.75	0.9207909	0.9591684	0.984003	● 0.9861142	0.9709395

The size-dependence of probabilities of large deviations for energy are shown in these tables. In this range of temperatures, for almost every  $u$ , the log-probability behaves noticeably in the same way. For small deviations, it

acts as linear function of  $L^{0.1}$  or  $L^{0.5}$  and for large deviations, the  $L^{0.5}$  or  $L^{1.0}$  dependence dominates with a really small difference in the R-squared values. Assuming that a large deviations principle holds for energy and coincide to its concentration bound, we deduce the concentration exponent for energy  $\rho = 2q = 1$  at sufficiently low temperature by the relation proved in [14].

### Concentration exponent of the correlation

The concentration estimate for empirical correlations follows from the analogous techniques simulation and from the empirical pair correlation result in [14].

$u$ \ T=1.50	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.013	0.9579837	0.986654	● 0.9996252	0.9911834	0.9673597
-0.012	0.956192	0.9854338	● 0.9993181	0.9919055	0.969125
-0.011	0.9651491	0.9905979	● 0.9995904	0.987495	0.9605089
-0.01	0.9638664	0.9899496	● 0.9997198	0.9883385	0.9619746
-0.005	0.9872596	● 0.9993066	0.9921019	0.9655708	0.9265579
0.007	0.9711726	0.9935154	● 0.9987168	0.9832262	0.9534157
0.008	0.9567664	0.9860005	● 0.9996266	0.991695	0.9681842
0.009	0.9476749	0.9805373	● 0.9987365	0.995099	0.9753938

$u$ \ T=2.00	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.07	0.941429	0.9751481	● 0.9952521	0.9942282	0.9775659
-0.06	0.9535059	0.9834213	● 0.9983730	0.9921847	0.9706888
-0.05	0.9631681	0.9894328	● 0.9995522	0.9886158	0.9627462
-0.01	0.9968279	● 0.9985737	0.9792138	0.9420948	0.8944357
-0.005	● 0.9996782	0.9943578	0.9667085	0.9224207	0.8689583
0.03	0.9609639	0.988227	● 0.9995925	0.9898018	0.9649148
0.04	0.9475861	0.9804369	● 0.9987193	0.9952427	0.9757364
0.05	0.9362537	0.9729062	0.9961561	● 0.9974934	0.9823594

$u$ \ T=2.24	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.5	0.9499842	0.9772916	● 0.9881121	0.9771421	0.9505674
-0.4	0.9765177	● 0.9836301	0.9693372	0.9352589	0.8887684
-0.3	0.9913134	● 0.9927382	0.9720664	0.9328405	0.8825777
-0.01	● 0.9996856	0.9892367	0.9559746	0.907232	0.8505173
-0.005	● 0.9993529	0.9874119	0.9525241	0.9024956	0.8448433
0.09	0.9800776	● 0.9972137	0.99583	0.9741677	0.9389354
0.1	0.9760887	0.9955789	● 0.9970681	0.9779979	0.9449574
0.15	0.9581648	0.9867022	● 0.9992893	0.9902613	0.9657128

$u$ \ T=2.26	$L^{0.1}$	$L^{0.5}$	$L^1$	$L^{1.5}$	$L^2$
-0.5	0.9565648	0.9817033	● 0.9908962	0.9796137	0.9539393
-0.4	0.9656315	0.985999	● 0.9894572	0.9730881	0.9431873
-0.01	● 0.9980849	0.9820646	0.9423537	0.8881065	0.8269874
-0.005	● 0.9976174	0.9807158	0.940031	0.8849999	0.8232974
0.18	0.9561101	0.983561	● 0.9953543	0.9861828	0.962064
0.2	0.9476055	0.9781697	● 0.9939331	0.9885157	0.9677283
0.22	0.944489	0.9765406	● 0.9940297	0.9900656	0.9704004

In this tables, we treat the concentration properties of the correlations between two pairs of spins at range  $x = 40$ . At  $T = 1.50$ , we observe mainly a surface-order dependence of the “asymptotic” exponential decay of the probability for sufficiently large deviations. For higher temperatures, the interval of “small” absolute deviations gets wider and wider and the log-probability tends to behave like  $L^{0.1}$  or  $L^{0.5}$ . Actually, this low order of growth emphasizes the phase transition phenomenon in which the correlation becomes more and more sensitive to the system and diverges at  $T_c$ . Such a simulation produces an optimal exponent for the correlation  $\rho = q = 1$  for lowest temperature and corresponds to the expected concentration result (see [14]).

## Concentration constants

In order to complete the study of stretched-exponential concentration, we illustrate the concentration constant behavior  $D_\rho$ . For each observable, we fix the concentration exponent  $\rho = 1$ , the large deviation  $u$  (for the magnetization, energy, correlation and suceptibility we choose respectively  $u = 0.007$ ,  $u = 0.075$ ,  $u = 0.009$  and  $u = 0.009$ ) and plot the corresponding curve according to the temperature (see 5.7).

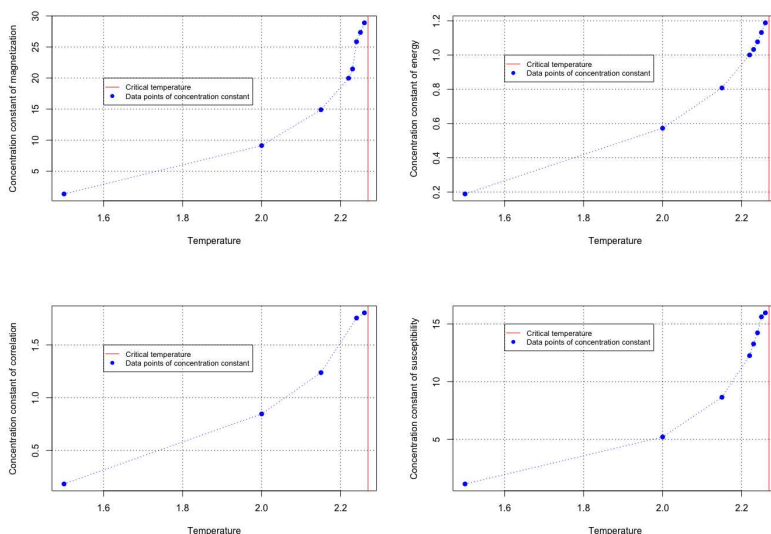


Figure 5.7: These plots represent the concentration constant for the magnetization (on the top left hand side), for the energy (on the top right hand side), for the correlation (on the bottom left hand side) and the susceptibility (on the bottom right hand side).

At this point, it is prudent to recall that these plots only give a qualitative representation of the concentration constants for one deviation  $u$  according to the temperature and the exponent. Even in this case, we observe a divergence of each concentration constant up to  $T_c$  which confirms the expected loss of the stretched-exponential bound at critical temperature (see 3.2.5).

## 5.4 Conclusion

In this chapter, we analyzed numerically the concentration parameters of the ferromagnetic Ising model in different temperature regimes. We were motivated by existing large deviation and concentration results which confront the correct asymptotic behavior of estimates with an upper bound for the logarithm of the exponential moment of all sufficiently smooth observables in every finite volume. In order to illustrate these results, we decided to measure observables from a configuration in a small size box inside an infinite (larger) one by simulating the Ising model with the classical Metropolis algorithm. This method produces accurate typical configurations according to the equilibrium state with a reasonable amount of time or Monte Carlo steps. Unfortunately, for rare events, this algorithm is not efficient and we tried without significant improvement a modified method to refine the study of concentration. Of course, this simple model can be more efficiently treated by different simulation strategies. In the high-temperature regime, we could relate the empirical Gaussian concentration constant to the variance of the concerned observable by fitting the histograms of fluctuations by a Gaussian density function at different temperature and compare their hypothetical divergence as power-laws. This result depicted the expected Gaussian concentration bound at very high temperature with the constant depending on the Dobrushin uniqueness condition and illustrated that a Gaussian concentration may hold until the critical temperature. Actually, these simulations motivated the study of concentration property in this range of temperatures in which we prove that Gaussian concentration bound for all  $\beta < \beta_c$ . At sufficiently low temperature, one expects a stretched-exponential concentration bound governed by a concentration exponent and a concentration constant. Unfortunately, this bound might be pessimistic and fitting histograms by the maximum likelihood density function in this case does not lead to the optimal stretched-exponential concentration parameters. We preferred calculating each temperature-dependent parameter independently starting by the concentration exponent. By fixing large-deviation events and the temperature, we computed the occurrences of fluctuations of estimates according to the size length of the box up to a concentration-related exponent and we deduced its value by a linear regression. At very low temperatures, we observed that the smallest exponent of all observables coincides with the stretched-exponential concentration bound. Close to the critical temperature, the smallest exponent drops down drastically and allows large deviations to occur more frequently. The second parameter describes qualitatively this behavior and shows the expected divergence occurring at critical temperature for each considered observables.

## 5.5 Code

To complete this computational chapter, we present the FORTRAN code we use to generate statistics on magnetization at different temperatures. Other observables can be simulated easily modifying the the subroutines mean and calculus.

```
1 program IsingModel
2
3
```

```

4      implicit none
5      ! Variables
6      integer :: i,j
7      character(len=64) :: filename
8      integer :: n=100
9      ! Spins
10     real, allocatable, dimension(:,:) :: M      ! Spin matrix
11     integer :: np = 10                          ! Number of spins in the small box
12     integer :: nMC, iMC                          ! Steps for the MC algorithm
13     integer :: neq, ieq                          ! Steps to reach the equilibrium
14     integer :: nmoy, imoy                         ! Number of averaging
15     integer :: ntps, itps                         ! Steps for different temperatures
16     integer :: a, amc
17     real :: right, left, up, down                ! Spin right/left/up/down
18     real :: x,y,z
19     real :: T, dT, maxT, minT                   ! Temperature
20     real :: dE                                   ! Difference of energy
21     real :: magn, ener                           ! Magnetization and energy
22     real :: moymagn, moyener                     ! Average magnetization and average energy
23     character(len=64) :: inputfile = 'input.in'
24     namelist /invars/ n, np, ntps, maxT, minT, neq, nmoy ;
25
26
27     open(UNIT=23, status='old', file=inputfile)
28     read(unit=23,NML=invars)
29     close(23)
30     write(*,invars)
31
32     call random_seed()
33
34
35     allocate (M(n+2,n+2))
36
37     !np=100
38     nMC=n*n
39     !ntps=2
40     !neq=100000
41     !nmoy=50000
42
43     !maxT=2.29
44     !minT=2.24
45     dT=(maxT-minT)/float(ntps-1)
46     T=minT
47
48
49     call ConfigInit()
50
51     do i=1,n+2
52     M(1,i)=1.0
53     M(i,1)=1.0
54     M(i,n+2)=1.0
55     M(n+2,i)=1.0
56     end do
57
58
59     do itps=1,ntps
60         !write(filename, '(a6,l3.3,a4)') 'T1=', itps, '.txt'
61         write(filename, '(a,F8.6,a)') 'T1=', T, '.txt'
62         open(unit=31, file=filename)
63         call equi()
64         call MC_step()
65         call mean()
66         call calculus()

```

```

67     close(31)
68
69     write(filename, '(a,F8.6,a)') 'MoyenmagnT1=', T, '.txt'
70     open(unit=32, file=filename, STATUS="replace", FORM="formatted", ACTION="write")
71     write(32,*) T, moymagn
72     close(32)
73
74     write(filename, '(a,F8.6,a)') 'M.T1=', T, '.txt'
75     open(unit=33, file=filename, STATUS="replace", FORM="unformatted", ACTION="write")
76     write(33) M
77     close(33)
78
79
80     T=T+dT
81 enddo
82
83
84
85 deallocate(M)
86
87
88 contains
89
90 ! Initial configuration
91
92 subroutine ConfigInit
93
94     call random_number(M)
95     do j=2,n+1
96         do i=2,n+1
97             if (M(i,j)>0.5) then
98                 M(i,j)=1.0
99             else
100                M(i,j)=-1.0
101            end if
102        end do
103    end do
104
105 end subroutine ConfigInit
106
107 ! Waiting for equilibrium
108
109 subroutine Equi
110
111     do ieq=1,neq
112         call MC_step()
113     enddo
114
115 end subroutine Equi
116
117 ! Averaging
118
119 subroutine mean
120     moymagn=0.0
121     do imoy=1,nmoy
122         magn=0.0
123         do a=1,2
124             call MC_step()
125         end do
126         do j=(n+2-np)/2,(n+2+np)/2
127             do i=(n+2-np)/2,(n+2+np)/2
128                 magn=magn+M(i,j)
129             end do

```

```

130         end do
131     moymagn=moymagn+magn/(float((np+1)*(np+1))*nmoy)
132     end do
133 end subroutine mean
134
135 ! Fluctuation calculus
136
137 subroutine calculus
138     do imoy=1,nmoy
139         magn=0.0
140         do a=1,2
141             call MC_step()
142         end do
143         do j=(n+2-np)/2,(n+2+np)/2
144             do i=(n+2-np)/2,(n+2+np)/2
145                 magn=magn+M(i,j)
146             end do
147         end do
148         print *, T, moymagn-magn/float((np+1)*(np+1))
149         write(31,*) moymagn-magn/float((np+1)*(np+1))
150     end do
151 end subroutine calculus
152
153 ! Monte Carlo method
154
155 subroutine MC_step
156
157     do iMC=1,nMC
158         call random_number(x)
159         i=floor(float(n-1)*x)+2
160         call random_number(y)
161         j=floor(float(n-1)*y)+2
162
163         right=i+1; left=i-1
164         up=j+1; down=j-1
165
166         dE=2*M(i,j)*(M(right,j)+M(left,j)+M(i,up)+M(i,down))
167         call random_number(z)
168         if (-dE/T>log(z)) M(i,j)=-M(i,j)
169
170     end do
171
172 end subroutine MC_step
173
174
175 end program IsingModel

```

## Chapter 6

# Gaussian concentration bound for the 2D Ising model

In the Dobrushin uniqueness regime, we recall that a Gaussian concentration bound holds for Gibbs measures associated to shift-invariant uniformly summable potentials on lattice systems. Here, motivated by the simulations in the previous chapter, we aim to prove that for the 2D ferromagnetic Ising model this result holds in the whole uniqueness regime except at the critical temperature when  $h = 0$  (see Remark 3.3). We can now state the main theorem of this chapter.

### Theorem 6.0.1.

1. For all  $\beta < \beta_c$  and  $h = 0$ , the 2D Ising model satisfies a Gaussian concentration bound.
2. For all  $\beta \geq 0$  and  $h > 0$ , the 2D Ising model satisfies a Gaussian concentration bound.

*Sketch of proof.* We sketch the proof which is based on several known results that we have to put together and we hope that there exists a more direct proof. We first use the fact that weak mixing implies strong mixing for all squares for 2D lattice spin systems (Section 6.1.1). Then we make use that strong mixing for all squares implies a logarithmic Sobolev inequality “LS” (Section 6.1.2) and we prove that a logarithmic Sobolev inequality implies a Gaussian concentration bound “GCB” (Section 6.1.3). Since the Ising model satisfies the weak mixing condition (for all  $\beta < \beta_c$  and  $h = 0$ ) or (for all  $\beta \geq 0$  and  $h > 0$ ), the theorem follows.  $\square$

## 6.1 Proof of Theorem 6.0.1

### 6.1.1 Weak mixing implies strong mixing for 2D lattice spin systems

In view of the proof of the main theorem, we need to define several notions of mixing, similar to the Dobrushin-Shlosman complete analyticity property, quantifying the sensitivity to boundary conditions. Here, we define the “weak”



and “strong” mixing properties of the Gibbs measure and mention their relation for 2D lattice spin systems. For more details, we refer the reader to [61].

Let  $\Lambda \in \mathbb{Z}^2$ . We recall the definition of total variation distance between two probability measures  $\mu$  and  $\nu$  on  $S^\Lambda$ :

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{\omega \in \Omega_\Lambda} |\mu(\omega) - \nu(\omega)| = \sup_{B \subset S^\Lambda} |\mu(B) - \nu(B)|. \quad (6.1.1)$$

This distance estimates the largest possible difference between two probability measures associated to the same event. The mixing criteria will be formulated in terms of maximal difference in expectation of a local observable, depending on the distance between the finite volume we consider and the support of a local perturbation on the boundary. To express this influence, we introduce the following measure

$$\gamma_{\Lambda, \Delta}^\Phi(\omega_\Delta | \eta) := \sum_{\omega \in \Omega_{\Lambda \setminus \Delta}} \gamma_\Lambda^\Phi(\omega | \eta) \quad (6.1.2)$$

where  $\Delta \subset \Lambda \in \mathbb{Z}^2$  and  $\Phi \in \mathcal{B}_R$ . We are now able to define two mixing conditions.

**Definition 6.1.1** (Strong and weak mixing).

Let  $\Phi \in \mathcal{B}_R$ . We say that  $\gamma^\Phi$  satisfies the “strong” mixing condition if there exist  $C, M > 0$  such that for all  $\Lambda \in \mathbb{Z}^d$ ,  $\Delta \subset \Lambda$  and for each  $x \in \Lambda^c$  we have

$$\sup_{\omega, \omega^{(x)} \in \Omega} \|\gamma_{\Lambda, \Delta}^\Phi(\cdot | \omega) - \gamma_{\Lambda, \Delta}^\Phi(\cdot | \omega^{(x)})\|_{TV} \leq C e^{-M d(\Delta, x)}. \quad (6.1.3)$$

We say that  $\gamma^\Phi$  satisfies the “weak” mixing condition if there exist  $C, M > 0$  such that for all  $\Lambda \in \mathbb{Z}^d$  and all  $\Delta \subset \Lambda$  we have

$$\sup_{\omega, \omega' \in S^{\Lambda^c}} \|\gamma_{\Lambda, \Delta}^\Phi(\cdot | \omega) - \gamma_{\Lambda, \Delta}^\Phi(\cdot | \omega')\|_{TV} \leq C \sum_{z \in \Delta, y \in \partial \Lambda} e^{-M \|z - y\|_\infty}. \quad (6.1.4)$$

Recall that  $\mathcal{B}_R$  is the set of potential of finite-range  $R > 0$ . The following theorem gives an equivalence between these two notions.

**Theorem 6.1.1** ([61]).

Let  $\Phi \in \mathcal{B}_R$ . In dimension two, the following are equivalent:

1. The Gibbs measures  $\gamma_\Lambda^\Phi$  satisfy the weak mixing condition for all  $\Lambda \in \mathbb{Z}^2$ .
2. The Gibbs measures  $\gamma_{\Lambda_n}^\Phi$  satisfy the strong mixing condition for all squares  $\Lambda_n$  with  $n > 0$ .

**Remark 6.1.** We note that the theorem ensures a strong mixing condition only for “regular” regions and may fail to arbitrarily shaped regions. This condition is also called *Restricted Complete Analyticity* and is stronger than weak mixing (see [4]). By regular regions, we mean surface built up from large enough squares. This argument restricts a boundary region to be one-dimensional and prevents “the boundary phase transition effect” to occur in

which the influence of a local perturbation of the boundary condition propagates in a half-space infinite-volume Gibbs measure only close to the boundary (see [60]).

**Remark 6.2.** We do not state the full version of Theorem 6.1.1. Regarding the proof, in the lattice  $\mathbb{Z}^d$  with  $d \geq 2$ , the authors showed by a sequence of intermediate results that strong mixing implies weak mixing (see Theorem 3.2 in [60]). Taking  $d = 2$  is necessary to ensure that weak mixing implies strong mixing.

In the high-temperature regime without external field, F. Martinelli, E. Olivieri and R.H. Schonmann also proved the following.

**Theorem 6.1.2** ([61]).

For the Ising model one has: for all  $\beta < \beta_c$  and  $h = 0$  there exist positive constants  $C, M > 0$  such that the Gibbs measures  $\gamma_\Lambda^\beta$  satisfy the weak mixing condition for all  $\Lambda \in \mathbb{Z}^2$ .

In fact, for the Ising model in arbitrary dimension  $d \geq 2$  one has: for all  $\beta < \beta_c$  there exist positive constants  $C, M > 0$  such that the Gibbs measures  $\gamma_\Lambda^\beta$  satisfy the weak mixing condition for all  $\Lambda \in \mathbb{Z}^d$  (see [61]).

Using the previous theorems, we deduce the following result for the Ising model.

**Corollary 6.1.1.**

For the Ising model one has: for all  $\beta < \beta_c$  and  $h = 0$  there exist positive constants  $C, M > 0$  such that the Gibbs measures  $\gamma_{\Lambda_n}^\beta$  satisfy the strong mixing condition for all squares  $\Lambda_n$  with  $n > 0$ .

For our purpose, we did not state the above result in its strongest version; namely that it remains valid for finite-range ferromagnetic Ising model in dimension two (in particular for all  $h > 0$ ) (see [61]) in which the potential  $\Phi \in \mathcal{B}_R$  is defined as follows.

$$\Phi(\Lambda, \omega) = \begin{cases} h\omega_x & \text{if } \Lambda = \{x\} \\ -J(x-y)\omega_x\omega_y & \text{if } \Lambda = \{x, y\} \\ 0 & \text{otherwise} \end{cases}$$

where  $J : \mathbb{Z}^2 \rightarrow \mathbb{R}$  is an even function such that  $J(0) = 0$ ,  $J(x) \geq 0$  if  $\|x\|_\infty \leq R$  and  $J(x) = 0$  if  $\|x\|_\infty > R$ .

In the whole uniqueness regime with a strictly positive external field, the authors of [77] generalized results of [61] concerning the low-temperature regime and proved the following.

**Theorem 6.1.3.** [77]

For the Ising model one has: for all  $\beta \geq 0$  and  $h > 0$  there exist positive constants  $C, M > 0$  such that the Gibbs measures  $\gamma_{\Lambda_n}^\beta$  satisfy the weak mixing condition for all squares  $\Lambda_n$  with  $n > 0$ .

We think it is worth to mention that for the ferromagnetic Ising model in arbitrary dimension, weak mixing implies that the pressure is analytic as a function of  $\beta$  and  $h$  (see [69]).

For the long-range case, these theorems may fail and we refer the reader to [25] for further details and examples.

### 6.1.2 Strong mixing implies the logarithmic Sobolev inequality

Let  $\mu$  be a probability measure on  $(\Omega, \mathfrak{B})$  and  $K : \Omega \rightarrow \mathbb{R}$  be a positive integrable function such that  $\int K |\log K| d\mu < +\infty$ . We define the entropy of  $K$  by

$$\text{Ent}_\mu(K) = \int K \log K d\mu - \int K d\mu \log \left( \int K d\mu \right).$$

For further details on this definition and consequences, we refer the reader to [1]. We now introduce the logarithmic Sobolev inequality.

**Definition 6.1.2.**

We say that  $\mu$  satisfies a logarithmic Sobolev inequality  $\text{LS}(D)$  if there exists  $D > 0$  such that for all  $K \in \mathcal{L}$ ,

$$\text{Ent}_\mu(K^2) \leq 2D \int \sum_{x \in \mathbb{Z}^2} \left| K(\omega) - K(\omega^{(x)}) \right|^2 d\mu(\omega). \quad (6.1.5)$$

For the sake of brevity, we say that  $\mu$  satisfies  $\text{LS}(D)$ . We define the logarithmic Sobolev constant  $D(\mu, \Lambda) < \infty$  as the smallest  $D$  such that  $\mu$  satisfies  $\text{LS}(D)$ .

An important aspect of logarithmic Sobolev inequalities is linked with the spectral gap properties of Markov stochastic dynamics converging to the equilibrium state. In fact, such estimates provide a relationship between mixing properties of the specification corresponding to the shift invariant, finite-range Gibbs potential  $\Phi$  and the unique Gibbs measure  $\mu_\Phi$  (see [80, 81]).

**Theorem 6.1.4 ([61]).**

Let  $\Phi \in \mathcal{B}_R$ . The following are equivalent:

1.  $\gamma_{\Lambda_n}^\Phi$  satisfies the strong mixing condition for all squares  $\Lambda_n$  with  $n > 0$ .
2.  $\gamma_{\Lambda_n}^\Phi$  satisfies  $\text{LS}(D)$  with finite constant  $D = \sup_{\Lambda_n \in \mathbb{Z}^d} D(\mu, \Lambda_n) < +\infty$ .

Here we only need a weaker version of this theorem.

**Corollary 6.1.2.**

Let  $\Phi \in \mathcal{B}_R$ . If  $\gamma_{\Lambda_n}^\Phi$  satisfies the strong mixing condition for all squares  $\Lambda_n$  with  $n > 0$  then  $\mu_\Phi$  satisfies  $\text{LS}(D)$  with finite constant  $\sup_{\Lambda_n \in \mathbb{Z}^d} D(\mu, \Lambda_n) < +\infty$ .

Therefore, the Ising model satisfies  $\text{LS}(D)$  for all  $\beta < \beta_c$  with the constant  $D$  depending only on  $\beta$  and the dimension.

### 6.1.3 Logarithmic Sobolev inequality implies GCB

The last step consists in connecting the logarithmic Sobolev inequality and the Gaussian concentration property. The implication relies on a famous method known as the Herbst argument that we adapt to our case (see [1, p. 252]). We stress that we are not dealing with Lipschitz functions but with local functions.

**Proposition 6.1.1.**

*If  $\mu$  satisfies LS( $D$ ), then  $\mu$  satisfies GCB( $D$ ).*

*Proof.* Let assume that  $K \in \mathcal{L}$ . We aim to apply the LS( $D$ ) property to the function  $\lambda \mapsto e^{\lambda K/2} \in \mathcal{L}$  for every  $\lambda \in \mathbb{R}_+$ . We introduce the function  $Z : [0, 1] \rightarrow \mathbb{R}$  and its derivative  $Z'$  defined by

$$Z(\lambda) = \int e^{\lambda K} d\mu \quad \text{and} \quad Z'(\lambda) = \int K e^{\lambda K} d\mu.$$

We emphasize the following useful bound whose proof is given in the appendix

**Lemma 6.1.1.**

*For all  $\lambda \in \mathbb{R}_+$  and all functions  $K : \Omega \rightarrow \mathbb{R}$ , we have*

$$\left| e^{\frac{\lambda}{2} K(\omega)} - e^{\frac{\lambda}{2} K(\omega^{(x)})} \right| \leq \frac{\lambda}{2} e^{\frac{\lambda}{2} \max(K(\omega); K(\omega^{(x)}))} \delta_x(K).$$

Therefore, by this lemma and by integration of the function  $\lambda \mapsto e^{\lambda K/2}$  over all configurations  $\omega \in \Omega$ , we can express (6.1.5) by a differential inequality involving the function  $Z$  and  $\lambda$  as follows

$$\lambda Z'(\lambda) \leq D \frac{\lambda^2}{2} \|\underline{\delta}(K)\|_2^2 Z(\lambda) + Z(\lambda) \log Z(\lambda). \quad (6.1.6)$$

The last step consists in integrating this inequality. To do so, we define  $g(\lambda) = \frac{1}{\lambda} \log Z(\lambda)$  with  $g(0) = \int_{\Omega} K d\mu$ . We observe that, for all  $\lambda > 0$

$$g'(\lambda) = -\frac{1}{\lambda^2} \log Z(\lambda) + \frac{Z'(\lambda)}{\lambda Z(\lambda)} \leq \frac{D}{2} \|\underline{\delta}(K)\|_2^2$$

which implies

$$g(\lambda) \leq \frac{D}{2} \|\underline{\delta}(K)\|_2^2 \lambda + g(0).$$

Taking  $\lambda = 1$  gives the expected Gaussian concentration bound (3.2.2).  $\square$

We can now conclude the proof of the Theorem 6.0.1. By Theorem 6.1.1, for two dimensional lattice systems and for all  $\Lambda \in \mathbb{Z}^2$ , the probability kernel  $\gamma_{\Lambda}^{\Phi}$  corresponding to the potential  $\Phi$  satisfying the weak mixing condition implies that it satisfies the strong mixing condition for all squares  $\Lambda_n$  with  $n > 0$ . Therefore, by Corollary 6.1.1 the specification associated to the 2-D Ising potential satisfies the strong mixing condition (for all  $\beta < \beta_c$  with  $h = 0$ ) or (for all  $\beta \geq 0$  with  $h > 0$ ) and all  $\Lambda_n$  with  $n > 0$ . By Theorem 6.1.4 the Ising model admits a unique Gibbs measure  $\mu_{\beta}$  which satisfies LS( $D$ ). By Proposition 6.1.1, it follows that for all (for all  $\beta < \beta_c$  with  $h = 0$ ) or (for all  $\beta \geq 0$  with  $h > 0$ ), the Ising model satisfies a Gaussian concentration bound.

### 6.1.4 Is Logarithmic Sobolev inequality equivalent to complete analyticity?

Let  $\eta$  be a fixed boundary condition. It is possible to integrate a measurable function  $K \in \mathcal{L}$  with respect to  $\gamma_\Lambda^\Phi(\cdot|\eta)$ . We denote by  $\gamma_\Lambda^\Phi K$  the  $\mathfrak{B}_\Lambda$ -measurable function defined by

$$\gamma_\Lambda^\Phi K(\eta) := \int K(\omega) \gamma_\Lambda^\Phi(d\omega|\eta). \quad (6.1.7)$$

In particular, one can compute the probability of any event in  $A \in \mathfrak{B}$  by summing over all the possible configurations and taking  $K$  as the indicator function of the event. This ingredient allows Dobrushin and Shlosman to introduce another mixing property equivalent to complete analyticity for  $\Phi$  (see [29]) and defined as follows.

#### Definition 6.1.3.

We say that  $\gamma^\Phi$  satisfies the Dobrushin-Shlosman mixing condition if there exists  $M > 0$  such that, for all  $\Lambda \in \mathbb{Z}^d$  there exists  $C > 0$  with the property that for all  $\Delta \subset \Lambda$ , for all  $x \in \partial\Delta$  and for all  $K \in \mathcal{L}$

$$\sup_{\eta, \eta^{(x)} \in \Omega} |\gamma_\Lambda^\Phi K(\eta) - \gamma_\Lambda^\Phi K(\eta^{(x)})| \leq C \|\delta(K)\|_1 e^{-Md(\Delta, x)}. \quad (6.1.8)$$

#### Lemma 6.1.2 ([81]).

Let  $\Phi \in \mathcal{B}_R$ . The following are equivalent:

1.  $\gamma_\Lambda^\Phi$  satisfies LS( $D$ ) with finite constant  $D = \sup_{\Lambda \in \mathbb{Z}^d} D(\mu, \Lambda) < +\infty$ .
2.  $\gamma_\Lambda^\Phi$  satisfies the Dobrushin-Shlosman mixing condition.

The argument to prove this lemma deals with the exponential decay of Markov semigroups for two different configurations (see Corollary 2.2 and Theorem 1.8 in [81]).

Finally, if complete analyticity holds for  $\Phi$  then the specification  $\gamma_\Lambda^\Phi$  satisfies a Gaussian concentration bound for all  $\Lambda \in \mathbb{Z}^d$ . For the Ising model, it seems that the reciprocal implication also holds. Beyond this case, we suggest that it may hold in any dimension and for any finite-range potential but we are not able to prove this at present. Precisely,

**CONJECTURE 1.** Let  $\Phi \in \mathcal{B}_R$ . The following are equivalent:

1.  $\gamma_\Lambda^\Phi$  satisfies GCB( $D$ ) with finite constant  $D = \sup_{\Lambda \in \mathbb{Z}^d} D(\mu, \Lambda) < +\infty$ .
2. Complete analyticity holds for  $\Phi$ .

## 6.2 Appendix: Proof of Lemma 6.1.1

We aim to prove that  $f_\lambda(a, b) = \lambda e^{\lambda \max(a, b)} |a - b| - |e^{\lambda a} - e^{\lambda b}|$  is positive for all  $\lambda \in \mathbb{R}_+$  and all  $a, b \in \mathbb{R}$ . Assume without loss of generalities that  $b < a$  then  $\max(a, b) = a$  and

$$f_\lambda(a, b) = \lambda e^{\lambda a} (a - b) - (e^{\lambda a} - e^{\lambda b}).$$

We set  $u = \lambda a$  and  $v = \lambda b$  then we obtain

$$f_1(u, v) = e^u(u - v - 1) + e^v.$$

Hence, by the classical inequality on the exponential one has

$$\begin{aligned}v - u + 1 &\leq e^{v-u} \\1 - e^{v-u} &\leq u - v \\e^u - e^v &\leq (u - v) e^u.\end{aligned}$$

Then  $f_\lambda(a, b) \geq 0$ . Finally, if we set  $a = \frac{\lambda}{2}K(\omega)$  and  $b = \frac{\lambda}{2}K(\omega^{(x)})$  for all  $\lambda \in \mathbb{R}_+$  we obtain the expected result.



## Chapter 7

# Gaussian concentration for potentials on $S^{\mathbb{N}}$ with subexponential variations

### 7.1 introduction

In this chapter, we study the concentration properties of equilibrium states on full shift spaces in dimension one associated with a potential with subexponential variations. Unlike the statistical mechanics point of view where the shift map is defined as the spatial translation on the lattice  $\mathbb{Z}^d$ , here we adopt the symbolic dynamics approach and consider the full-shift map as a temporal evolution of all possible sequences of symbols. Such discrete-time processes are known as chains with complete connections in the theory of stochastic processes. Our concentration result describes the fluctuations of observables of the form  $K(x, Tx, \dots, T^{n-1}x)$  around their average. The only restriction on  $K$  is that it has to be separately Lipschitz. By this we mean that, for all  $i = 0, \dots, n-1$ , there exists a constant  $\text{Lip}_i(K)$  with

$$|K(x_0, \dots, x_i, \dots, x_{n-1}) - K(x_0, \dots, x'_i, \dots, x_{n-1})| \leq \text{Lip}_i(K) d(x_i, x'_i).$$

for all points  $x_0, \dots, x_i, \dots, x_{n-1}, x'_i$  in  $\Omega$ , where  $d$  is the usual distance on  $\Omega$  (see (7.2.1)). So  $K$  can be nonlinear and implicitly defined. Of course, such a class contains partial sums of Lipschitz functions, namely functions of the form  $K(x_0, \dots, x_{n-1}) = f(x_0) + \dots + f(x_{n-1})$  for which  $\text{Lip}_i(K) = \text{Lip}(f)$  for all  $i$ . Beside considering very general observables, the other essential characteristics of concentration inequalities is that they are valid for all  $n$ . More precisely, we shall prove the following “Gaussian concentration bound”. There exists a constant  $C$  such that, for all  $n$  and for all separately Lipschitz functions  $K(x_0, \dots, x_{n-1})$ , we have

$$\begin{aligned} & \int \exp\left(K\left(x, Tx, \dots, T^{n-1}x\right)\right) d\mu_\phi(x) \\ & \leq \exp\left(\int K\left(x, Tx, \dots, T^{n-1}x\right) d\mu_\phi(x)\right) \exp\left(C \sum_{j=0}^{n-1} \text{Lip}_j(K)^2\right). \end{aligned} \quad (7.1.1)$$



The crucial point is that  $C$  is independent of  $n$  and  $K$ . By a standard argument (see below), the previous inequality implies that for all  $u > 0$

$$\begin{aligned} & \mu_\phi \left( x : K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu_\phi(y) \geq u \right) \\ & \leq \exp \left( -\frac{u^2}{4C \sum_{i=0}^{n-1} \text{Lip}_i(K)^2} \right). \end{aligned} \quad (7.1.2)$$

We insist on the fact that we consider functions  $K$  on  $\Omega^n$  with a classical Lipschitz condition different of the  $\delta$ 's condition (3.3.2). The Gaussian concentration bound (7.1.1) is known for Lipschitz potentials [17]. We shall prove that it remains true for a large subclass of potentials  $\phi$  satisfying Walters condition. For instance, the bound holds for a potential whose variation is  $O(n^{-\alpha})$  for some  $\alpha > 2$ . Of course, different types of concentration inequalities have been obtained for more general potentials, and for more general ‘‘chaotic’’ dynamical systems, in particular the authors in [17] obtained different types of concentration properties for nonuniformly hyperbolic dynamical systems modeled by Young towers.

We apply the Gaussian concentration bound and its consequences, like (7.1.2), to various observables. On the one hand, we obtain concentration bounds for previously studied observables. We get the same bounds but they are no more limited to equilibrium states with Lipschitz potentials. On the other hand, we consider observables not considered before. Even when  $K(x, \dots, T^{n-1}x) = S_n f(x)$ , we get a non-trivial bound. We then obtain a control on the fluctuations of the empirical frequency of blocks  $a^0, \dots, a^{k-1}$  around  $\mu([a^0, \dots, a^{k-1}])$ , uniformly in  $a^0, \dots, a^{k-1} \in A^k$ . We then consider an estimator of the entropy  $\mu_\phi$  based on hitting times. The next application is about the speed of convergence of the empirical measure  $(1/n) \sum_{i=0}^{n-1} \delta_{T^i x}$  towards  $\mu_\phi$  in Wasserstein distance. Then we obtain an upper bound for the  $\bar{d}$ -distance between any shift-invariant probability measure and  $\mu_\phi$ . This distance is bounded by the square root of their relative entropy, times a constant. A consequence of this inequality is a bound for the speed of convergence of the Markov approximation of  $\mu_\phi$  in  $\bar{d}$ -distance. Then we quantify the ‘‘shadowing’’ of an orbit by another one which has to start in a subset of  $\Omega$  with  $\mu_\phi$ -measure  $1/3$ , say. Finally, we prove an almost-sure version of the central limit theorem. This application shows in particular that concentration inequalities can also be used to obtain limit theorems. This result may also hold for the more general case of subshifts of finite type which are described by a finite list of forbidden patterns or words. The result was published in [20] and follows from the speed of convergence of the iterates of the Ruelle-Perron-Frobenius operator to the unique equilibrium state [66] in  $\text{sup}$ -norm. In this thesis, a minor mistake has been corrected in the estimation of the Lipschitz-norm.

## 7.2 Settings

Let  $\Omega = S^{\mathbb{N}}$  where  $S$  is a finite set so that  $\Omega$  is a compact set. Let  $\mathcal{M}(\Omega)$  be the set of all probability measures on the Borel  $\sigma$ -algebra  $\mathfrak{B}$  of  $\Omega$ . This set is convex and compact in the weak topology. We denote by  $x = x^0 x^1 \dots$  the elements of  $\Omega$  (hence  $x^i \in S$ ), and by  $T$  the shift map:  $(Tx)^k = x^{k+1}$ ,  $k \in \mathbb{N}$ . (We use upper indices instead of lower indices because we will need

to consider bunches of points in  $\Omega$ , e.g.,  $x_0, x_1, \dots, x_p, x_i \in \Omega$ .) Recall that we use the classical distance

$$d_\theta(x, y) = \theta^{\inf\{k: x^k \neq y^k\}} \quad (7.2.1)$$

where  $\theta \in (0, 1)$  is some fixed number. Probability measures are defined on the Borel sigma-algebra of  $\Omega$  which is generated by cylinder sets.

**Definition 7.2.1.**

We say that  $\phi : \Omega \rightarrow \mathbb{R}$  is a continuous potential, if

$$\text{var}_n(\phi) := \sup\{|\phi(x) - \phi(y)| : x^i = y^i, 0 \leq i \leq n-1\} \xrightarrow{n \rightarrow \infty} 0.$$

The sequence  $(\text{var}_n(\phi))_{n \geq 1}$  is the modulus of continuity of  $\phi$  and it is called the 'variation' of  $\phi$ .

By the way, we denote by  $\mathcal{C}(\Omega)$  the Banach space of real-valued continuous functions on  $\Omega$  equipped with the supremum norm  $\|\cdot\|_\infty$ . Let  $\phi \in \mathcal{C}(\Omega)$ . We put further restrictions on  $\phi$ , namely that it must satisfy the Walters condition [84]. For  $x, y$  in  $\Omega$  let

$$W(\phi, x, y) = \sup_{n \in \mathbb{N}} \sup_{a \in A^n} |S_n \phi(ax) - S_n \phi(ay)|.$$

We assume that  $W(\phi, x, y)$  exists and that there exists  $W(\phi) > 0$  such that

$$\sup_{x, y \in \Omega} W(\phi, x, y) \leq W(\phi). \quad (7.2.2)$$

Now for  $p \in \mathbb{N}$  let

$$W_p(\phi) := \sup\{W(\phi, x, y) : x^i = y^i, 0 \leq i \leq p-1\}.$$

**Definition 7.2.2.**

$\phi$  is said to satisfy Walters' condition if  $(W_p(\phi))_{p \in \mathbb{N}}$  is a strictly positive sequence and decreases to 0 as  $p \rightarrow \infty$ .

We now make several remarks on Walters' condition. First, observe that locally constant potentials do not satisfy this condition because  $W_p(\phi) = 0$  for all  $p$  larger than some  $p_0$ . But one can in fact work with any strictly positive sequence  $(\widetilde{W}_p(\phi))_{p \in \mathbb{N}}$  decreasing to zero such that  $W_p(\phi) \leq \widetilde{W}_p(\phi)$  for all  $p$ , e.g.,  $\max(W_p(\phi), \eta^p)$  for some fixed  $\eta \in (0, 1)$ . Second, one easily checks that

$$\text{var}_{p+1}(\phi) \leq W_p(\phi) \leq \sum_{k=p+1}^{\infty} \text{var}_k(\phi), \quad p \in \mathbb{N}. \quad (7.2.3)$$

Hence the set of potentials satisfying Walters' condition contains the set of potentials with summable variation. In particular,  $(W_p(\phi))_p$  is bounded above by a geometric sequence if and only if  $(\text{var}_p(\phi))_p$  is also bounded above by a geometric sequence. This corresponds to the case of Lipschitz or Hölder potentials (with respect to  $d_\theta$ ).

Now define the transfer operator  $P_\phi$  or commonly Ruelle's Perron-Frobenius operator  $P_\phi : \mathcal{C}(\Omega) \rightarrow \mathcal{C}(\Omega)$  as

$$P_\phi f(x) = \sum_{Ty=x} f(y) e^{\phi(y)}$$

and we denote by  $P_\phi^* : \mathcal{M}(\Omega) \rightarrow \mathcal{M}(\Omega)$  the dual operator of  $P_\phi$  as

$$\forall f \in \mathcal{C}(\Omega), \quad \int P_\phi f d\mu = \int f d(P_\phi^* \mu). \quad (7.2.4)$$

The next step is to define a function space preserved by  $P_\phi$  and on which it has good spectral properties. The study of such operators is important because it relates its spectral properties to the measure (see [54]). We take the space of Lipschitz functions with respect to a new distance  $d_\phi$  built out of  $\phi$  as follows.

**Definition 7.2.3** (The distance  $d_\phi$ ).

For  $x, y \in \Omega$  let

$$d_\phi(x, y) = W_p(\phi) \quad \text{if} \quad d_\theta(x, y) = \theta^p$$

and  $d_\phi(x, x) = 0$ .

Now define

$$\mathcal{L}_\phi = \{f \in \mathcal{C}(\Omega) : \exists M > 0 \text{ such that } \text{var}_n(f) \leq MW_n(\phi), n = 1, 2, \dots\}$$

and

$$\text{Lip}_\phi(f) = \sup \left\{ \frac{|f(x) - f(y)|}{d_\phi(x, y)} : x \neq y \right\} = \sup \left\{ \frac{\text{var}_n(f)}{W_n(\phi)} : n \in \mathbb{N} \right\}.$$

One can then define a norm on  $\mathcal{L}_\phi$ , making it a Banach space, by setting

$$\|f\|_{\mathcal{L}_\phi} = \|f\|_\infty + \text{Lip}_\phi(f).$$

**Remark 7.1.** The usual Banach space of Lipschitz functions is defined as follows. Let

$$\mathcal{L}_\theta = \{f \in \mathcal{C}(\Omega) : \exists M > 0 \text{ such that } \text{var}_n(f) \leq M\theta^n, n = 1, 2, \dots\}$$

and

$$\text{Lip}_\theta(f) = \sup \left\{ \frac{|f(x) - f(y)|}{d_\theta(x, y)} : x \neq y \right\} = \sup \left\{ \frac{\text{var}_n(f)}{\theta^n} : n \in \mathbb{N} \right\}.$$

The canonical norm making  $\mathcal{L}_\theta$  a Banach space is  $\|f\|_{\mathcal{L}_\theta} = \|f\|_\infty + \text{Lip}_\theta(f)$ .

In view of (7.2.3), if we have  $W_n(\phi) = O(\theta^n)$ , then  $\mathcal{L}_\theta = \mathcal{L}_\phi$ . If we now have, for instance,  $W_n(\phi) = O(n^{-q})$  for some  $q > 0$ , then we get a bigger space which contains in particular all functions  $f$  such that  $\text{var}_n(f) = O(n^{-r})$  with  $r \geq q$ .

For the one-sided subshift of finite type associated to a Lipschitz potential, P. Walters and D. Ruelle proved in [83, 11, 72] the quasi-compactness of the transfer operator. Imposing to the subshift of finite type to be aperiodic it follows that the operator has a spectral gap property, i.e., there exists a gap between the first eigenvalue and the second one of the transfer operator. For more details on spectral theory of quasi-compact operators, we refer the reader to [41, 2]. This gap implies the existence of an exponentially mixing unique equilibrium state which satisfies a Gaussian concentration bound

(see [17]). This property is derived from the estimation of the speed of convergence of the transfer operator. Once the condition of quasi-compactness is lost, cones method and projective metrics introduced by Birkhoff would estimate the sub-exponential decay of correlation depending on the contraction coefficient of cones for a given considered potential. In her thesis, V. Maume gave such estimation for potential with subexponential variations and proved that the central limit theorem holds by using relevant set of cones. Up to this point, one may ask for conditions about the unique equilibrium state associated to such potential in which the Gaussian concentration bound holds. To do so, a classical recipe is implemented and relies on two main ingredients. The first is to decompose the difference between the observable and its mean as a telescopic sum of martingale differences and the second uses the spectral gap of the Ruelle-Perron-Frobenius operator in order to obtain a uniform estimate of the martingale difference. Then, we use the classical Azuma-Hoeffding inequality 7.5.7 to prove that the measure satisfies the Gaussian concentration bound.

### 7.3 Speed of convergence of the transfer operator

The following result is instrumental to this chapter. In brief, it tells us that a potential  $\phi$  satisfying Walters' condition has a unique equilibrium state, which will be denoted by  $\mu_\phi$ , and gives a speed of convergence for the properly normalized iterates of the associated Ruelle's Perron-Frobenius operator acting on an initial Lipschitz function. The first part of the theorem is due to Walters, while the second one is due to Maume-Deschamps and can be found in her PhD thesis in french [66, Chapter I.2]. Unfortunately, her result was not published even though it is much sharper than the result in [46]. Hence, for completeness we provide the proof in appendix.

**Theorem 7.3.1** ([84], [66]).

Let  $\phi : \Omega \rightarrow \mathbb{R}$  satisfying Walters' condition as above. Then the following holds.

- A. There exists a unique triplet  $(h_\phi, \lambda_\phi, \nu_\phi)$  such that  $h_\phi \in \mathcal{L}_\phi$  and is strictly positive,  $\|\log h_\phi\|_\infty < \infty$ ,  $\lambda_\phi > 0$ ,  $\nu_\phi$  a fully supported probability measure such that  $\int h_\phi d\nu_\phi = 1$ . Moreover,  $P_\phi h_\phi = \lambda_\phi h_\phi$  and  $P_\phi^* \nu_\phi = \lambda_\phi \nu_\phi$ , and  $\phi$  has a unique equilibrium state  $\mu_\phi = h_\phi \nu_\phi$  which is mixing.
- B. There exists a positive sequence  $(\epsilon_n)_{n \in \mathbb{N}}$  converging to zero, such that, for any  $f \in \mathcal{L}_\phi$ ,

$$\left\| \frac{P_\phi^n f}{\lambda_\phi^n} - h_\phi \int f d\nu_\phi \right\|_\infty \leq C_{(7.3.1)} \epsilon_n \|f\|_{\mathcal{L}_\phi}, \quad \forall n \in \mathbb{N}. \quad (7.3.1)$$

Moreover, one has the following behaviors:

- 1. If  $W_n(\phi) = O(\eta^n)$  for some  $\eta \in (0, 1)$ , then there exists  $\eta' \in (0, 1)$  such that  $\epsilon_n = O(\eta'^n)$ .
- 2. If  $W_n(\phi) = O(n^{-\alpha})$  for some  $\alpha > 0$ , then  $\epsilon_n = O(n^{-\alpha})$ .

3. If  $W_n(\phi) = O(\theta^{(\log n)^\alpha})$  for some  $\theta \in (0, 1)$  and  $\alpha > 1$ , then, for any  $\epsilon > 0$ ,  $\epsilon_n = O(\theta^{(\log n)^{\alpha-\epsilon}})$ .
4. If  $W_n(\phi) = O(e^{-cn^\alpha})$  for some  $c > 0$  and  $\alpha \in (0, 1)$ , then there exists  $c' > 0$  such that  $\epsilon_n = O(e^{-c'n^{\frac{\alpha}{\alpha+1}}})$ .

We recall that  $\mu_\phi$  is an equilibrium state means that it maximizes the functional  $\mu \mapsto h(\nu) + \int \phi d\nu$  over the set of shift-invariant probability measures on  $\Omega$ , where  $h(\nu)$  is the entropy of  $\nu$ , and the maximum is equal to the topological pressure  $P(\phi)$  of  $\phi$  (see e.g. [44]), and we have  $P(\phi) = \log \lambda_\phi$ .

The first behavior corresponds to a geometric or Hölderian potential with respect to the usual distance and implies an exponential speed of convergence to the equilibrium state and exponential mixing properties.

Let us give examples of potentials. First consider  $A = \{-1, 1\}$  and  $p > 1$ , and define

$$\phi(x) = - \sum_{n \geq 2} \frac{x^0 x^{n-1}}{n^p}.$$

One can check that  $W_n(\phi) = O(n^{-p+2})$ . This is the analogue of the so-called long-range Ising model on  $\mathbb{N}$ . Let us now take  $A = \{0, 1\}$  and let  $[0^k 1] = \{x \in \Omega : x^i = 0, 0 \leq i \leq k-1, \text{ and } x_k = 1\}$ . Let  $(v_n)$  be a monotone decreasing sequence of real numbers converging to 0 and define

$$\phi(x) = \begin{cases} v_k & \text{if } x \in [0^k 1] \\ 0 & \text{if } x = (0, 0, \dots) \\ 0 & \text{otherwise.} \end{cases}$$

One can check that  $\text{var}_n(\phi) = v_n$ . This example is taken from [71].

**Remark 7.2.** *Let us briefly explain how we can interpret an equilibrium state for a non-Lipschitz potential as an absolutely continuous invariant measure of a piecewise expanding map of the unit interval with a Markov partition. It is well-known that a uniformly expanding map  $S$  of the unit interval with a finite Markov partition which is piecewise  $C^{1+\eta}$ , for some  $\eta > 0$ , can be coded by a subshift of finite type  $(\Omega, T)$  over a finite alphabet. Then,  $-\log |S'|$  induces a potential  $\phi$  on  $\Omega$  which is Lipschitz (with respect to  $d_\theta$ ). The pullback of  $\mu_\phi$  is then the unique absolutely continuous invariant probability measure for  $S$ . In [23], the authors showed that, given  $\phi$  which is not Lipschitz, one can construct a uniformly expanding map of the unit interval with a finite Markov partition which is piecewise  $C^1$ , but not piecewise  $C^{1+\eta}$  for any  $\eta > 0$ , and such that the pullback of  $\mu_\phi$  is the Lebesgue measure.*

## 7.4 Gaussian concentration bound

In this sub-section we prove the following result:

### Theorem 7.4.1.

Suppose that  $\phi$  satisfies one of the following conditions:

1.  $W_n(\phi) = O(\theta^n)$  (that is,  $\phi$  is  $d_\theta$ -Lipschitz);
2.  $W_n(\phi) = O(n^{-\alpha})$  for some  $\alpha > 1$ ;
3.  $W_n(\phi) = O(\theta^{(\log n)^\alpha})$  for some  $\theta \in (0, 1)$  and  $\alpha > 1$ ;
4.  $W_n(\phi) = O(e^{-cn^\alpha})$  for some  $c > 0$  and  $\alpha \in (0, 1)$ .

Then the process  $(x, Tx, \dots)$ , with  $x$  distributed according to  $\mu_\phi$ , satisfies the following Gaussian concentration bound. There exists  $C_{(7.4.1)} > 0$  such that for any  $n \in \mathbb{N}$  and for any separately  $d_\theta$ -Lipschitz function  $K : \Omega^n \rightarrow \mathbb{R}$ , we have

$$\begin{aligned} & \int \exp\left(K\left(x, Tx, \dots, T^{n-1}x\right)\right) d\mu_\phi(x) \\ & \leq \exp\left(\int K\left(x, Tx, \dots, T^{n-1}x\right) d\mu_\phi(x)\right) \exp\left(C_{(7.4.1)} \sum_{j=0}^{n-1} \text{Lip}_{\theta,j}(K)^2\right). \end{aligned} \tag{7.4.1}$$

Three remarks are in order. First, we conjecture that this theorem is valid under the condition  $\sum_n \text{var}_n(\phi) < \infty$ . Second, it would be useful to have an explicit formula for  $C_{(7.4.1)}$  in (7.4.1). Unfortunately, this constant is proportional to  $C_{(7.3.1)}$  (see Theorem 7.3.1) which is cumbersome since it involves the eigendata of  $P_\phi$ . Third, for the sake of simplicity, we considered the full shift  $S^{\mathbb{N}}$ . In fact, our results remain true if  $\Omega \subset S^{\mathbb{N}}$  is a topologically mixing one-sided subshift of finite type. Moreover, one can extend Theorem 7.4.1 to bilateral subshifts of finite type by a trick used in [17].

We now give some corollaries of our main theorem that we will be used in the section on applications. First, by (7.2.3) we immediately obtain the following corollary.

### Corollary 7.4.1.

If there exists  $\alpha > 2$  such that

$$\text{var}_n(\phi) = O\left(\frac{1}{n^\alpha}\right)$$

then we have the Gaussian concentration bound (7.4.1).

Next, we get the following concentration inequalities from (7.4.1).

**Corollary 7.4.2.**

For all  $u > 0$ , we have

$$\begin{aligned} & \mu_\phi \left( x \in \Omega : K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu_\phi(y) \geq u \right) \\ & \leq \exp \left( -\frac{u^2}{4C_{(7.4.1)} \sum_{i=0}^{n-1} \text{Lip}_{\theta,i}(K)^2} \right) \end{aligned} \quad (7.4.2)$$

and

$$\begin{aligned} & \mu_\phi \left( x \in \Omega : \left| K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu_\phi(y) \right| \geq u \right) \\ & \leq 2 \exp \left( -\frac{u^2}{4C_{(7.4.1)} \sum_{i=0}^{n-1} \text{Lip}_{\theta,i}(K)^2} \right). \end{aligned} \quad (7.4.3)$$

*Proof.* Inequality (7.4.2) follows by Chernoff's bounding method 3.1.1. Let us give the proof for completeness. Let  $u > 0$ . For any random variable  $Y$ , Markov's inequality tells us that  $\mathbb{P}(Y \geq u) \leq e^{-\xi u} \mathbb{E}(e^{\xi Y})$  for all  $\xi > 0$ . Now let

$$Y = K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu_\phi(y).$$

Using (7.4.1) and optimizing over  $\xi$ , we get (7.4.2). Inequality (7.4.3) follows by applying (7.4.2) to  $-K$  and then summing up the two bounds.  $\square$

The last corollary we want to state is about the variance of any separately  $d_\theta$ -Lipschitz function.

**Corollary 7.4.3.**

We have

$$\begin{aligned} & \int \left( K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu_\phi(y) \right)^2 d\mu_\phi(x) \\ & \leq 2C_{(7.4.1)}^2 \sum_{i=0}^{n-1} \text{Lip}_{\theta,i}(K)^2. \end{aligned} \quad (7.4.4)$$

*Proof.* To alleviate notations, we simply write  $K$  instead of  $K(x, Tx, \dots, T^{n-1}x)$ ,  $\int K$  instead of  $\int K(y, Ty, \dots, T^{n-1}y) d\mu_\phi(y)$ , and so on and so forth. Applying (7.4.1) to  $\xi K$  where  $\xi$  is any real number different from 0, we get

$$\int \exp \left( \xi \left( K - \int K \right) \right) \leq \exp \left( C_{(7.4.1)} \xi^2 \sum_{j=0}^{n-1} \text{Lip}_{\theta,j}(K)^2 \right).$$

Now by Taylor expansion we get

$$1 + \frac{\xi^2}{2} \int \left( K - \int K \right)^2 + o(\xi^2) \leq 1 + C_{(7.4.1)} \xi^2 \sum_{j=0}^{n-1} \text{Lip}_{\theta,j}(K)^2 + o(\xi^2).$$

Dividing by  $\xi^2$  on both sides and then taking the limit  $\xi \rightarrow 0$ , we obtain the desired inequality.  $\square$

Although we were not able to prove the Gaussian concentration bound for separately  $d_\phi$ -Lipschitz functions, for many applications separately  $d_\phi$ -Lipschitz functions are more natural. Furthermore there is a notable class of separately  $d_\phi$ -Lipschitz functions, namely Birkhoff sums of the potential itself, for which our theorem holds. Actually, when  $\phi \in \mathcal{L}_\phi$ , the function  $K(x, \dots, T^{n-1}x) = S_n\phi(x)$  is obviously separately  $d_\phi$ -Lipschitz and  $\text{Lip}_{\phi,j}(K) = \text{Lip}_\phi(\phi)$  for all  $j$ . We have the following result.

**Theorem 7.4.2.** *Under the hypotheses of Theorem 7.4.1, there exists  $C_{(7.4.2)} > 0$  such that, for any  $\psi \in \mathcal{L}_\phi$ , for all  $u > 0$ , and for all  $n \in \mathbb{N}$ , we have*

$$\begin{aligned} & \mu_\phi \left( x \in \Omega : \frac{1}{n} S_n \psi(x) - \int \psi \, d\mu_\phi \geq u \right) \\ & \leq \exp \left( - \frac{nu^2}{4C_{(7.4.2)} \text{Lip}_\phi(\psi)^2} \right). \end{aligned} \quad (7.4.5)$$

The proof is left to the reader. The main (simple) modification lies in the proof of Lemma 7.5.3 in which considering a Birkhoff sum of a  $d_\phi$ -Lipschitz function works fine, whereas we are stuck for a general separately  $d_\phi$ -Lipschitz function.

We will apply this result with  $\psi = -\phi$  to derive concentration bounds for hitting times. Note that under the assumptions of this theorem,  $\{\psi(T^n x)\}_{n \geq 0}$  satisfies the central limit theorem [66, Chapter 2].

## 7.4.1 Related works

The novelty here is to prove a Gaussian concentration bound for potentials with a variation decaying subexponentially. For  $\phi$  is Lipschitz, Theorem 7.4.1 was proved in [17]. The main goal of [17] was then to deal with nonuniformly hyperbolic systems modeled by a Young tower. For a tower with a return-time to the base with exponential tails, the authors of [17] proved a Gaussian concentration bound. For polynomial tails, they proved moment concentration bounds. For  $C^{1+n}$  maps of the unit interval with an indifferent fixed point, which are thus nonuniformly expanding, we are in the latter situation. In view of Remark 7.2 above, we deal here with maps whose derivative is not Hölder continuous, but which are still uniformly expanding.

Let us also recall that the authors in [38] prove a Gaussian concentration bound for  $\phi$  of summable variation (whereas we need a bit more than summable). Their proof is based on coupling. However, they consider functions  $K$  on  $S^n$ , not on  $(S^{\mathbb{N}})^n = \Omega^n$  as in our case. For such functions, the analogue of  $\text{Lip}_{\theta,i}(K)$  is

$$\delta_i(K) = \sup\{|K(a^0, \dots, a^i, \dots, a^{n-1}) - K(b^0, \dots, b^i, \dots, b^{n-1})| : a^j = b^j, \forall j \neq i\}.$$

It is clear that a Gaussian concentration bound for functions  $K : (S^{\mathbb{N}})^n \rightarrow \mathbb{R}$  implies a Gaussian concentration bound for functions  $K : S^n \rightarrow \mathbb{R}$ , but the converse is not true. We also recall that a Gaussian concentration bound holds in the more general case of chains with complete connections on countable alphabets (see [16]).



## 7.5 Proof of Theorem 7.4.1

We follow the proof given in [17] with the appropriate modifications to go beyond Lipschitz potentials.

### 7.5.1 Some preparatory results

It is convenient to normalize the potential  $\phi$  or, equivalently, the operator  $P_\phi$  in the following way. We use the notations of Theorem 7.3.1. Let

$$\tilde{P}_\phi f = \lambda_\phi^{-1} h_\phi^{-1} P_\phi(f h_\phi).$$

Thus

$$\tilde{P}_\phi 1 = 1 \quad \text{and} \quad \tilde{P}_\phi^* \mu_\phi = \mu_\phi. \quad (7.5.1)$$

Let  $g$  denote the inverse of the Jacobian of  $T$ , and  $g^{(k)}$  the inverse of the Jacobian of  $T^k$ , that is,

$$g = \frac{h_\phi}{\lambda_\phi h_\phi \circ T} \exp(\phi) \quad \text{and} \quad g^{(k)} = \frac{h_\phi}{\lambda_\phi^k h_\phi \circ T^k} \exp\left(\sum_{i=0}^{k-1} \phi \circ T^i\right).$$

(Of course  $g = g^{(1)}$ .) Therefore we have

$$\tilde{P}_\phi^n f(x) = \sum_{Ty=x} g(y) f(y) \quad \text{and} \quad \tilde{P}_\phi^k f(x) = \sum_{T^k y=x} g^{(k)}(y) f(y). \quad (7.5.2)$$

Estimate (7.3.1) now takes the form

$$\left\| \tilde{P}_\phi^n f - \int f d\mu_\phi \right\|_\infty \leq C_{(7.3.1)} \|f\|_{\mathcal{L}_\phi} \epsilon_n, \quad n \geq 1, \quad (7.5.3)$$

for any  $f \in \mathcal{L}_\phi$ . Finally, we will need the following distortion estimate. Let  $x, y \in \Omega$  such that  $x^i = y^i$  for  $i = 0, \dots, n-1$  and  $x', y' \in \Omega$  such that  $T^k x' = x$  and  $T^k y' = y$ . Then it is easy to check (see [66, Chapter 2]) that, for any  $k$ ,

$$\left| 1 - \frac{g^{(k)}(x')}{g^{(k)}(y')} \right| \leq c_{(7.5.4)} d_\phi(x, y) \quad (7.5.4)$$

for some constant  $c_{(7.5.4)} > 0$  depending only on  $\phi$ .

We will use the following inequality relating the distances  $d_\theta$  and  $d_\phi$ .

#### Lemma 7.5.1.

Suppose that  $W_n(\phi) = O(\theta^n)$ ,  $n \geq 1$ , or

$$\lim_n \frac{W_n(\phi)}{W_{n+1}(\phi)} = 1. \quad (7.5.5)$$

Then there exists  $c_{(7.5.1)} > 0$

$$\sup_n \frac{\theta^n}{W_n(\phi)} \leq c_{(7.5.1)}$$

or, equivalently,

$$d_\theta(x, y) \leq c_{(7.5.1)} d_\phi(x, y)$$

for all  $x, y$ .

*Proof.* The statement is trivial when  $W_n(\phi) = O(\theta^n)$ . if (7.5.5) holds, then there exists  $n_0$  such that for all  $n \geq n_0$

$$\frac{W_n(\phi)}{W_{n+1}(\phi)} \leq \frac{1}{\theta},$$

hence  $W_n(\phi) \geq \theta^{n-n_0} W_{n_0}(\phi)$ . Then the desired inequalities follow easily from the definitions.  $\square$

## 7.5.2 Proof of Theorem 7.4.1

Fix a separately  $d_\theta$ -Lipschitz function  $K : \Omega^n \rightarrow \mathbb{R}$ . It is convenient to think of it as a function on  $\Omega^{\mathbb{N}}$  depending only on the first  $n$  coordinates, therefore  $\text{Lip}_{\theta,i}(K) = 0$  for  $i \geq n$ . We endow  $\Omega^{\mathbb{N}}$  with the measure  $\mu^\infty$  obtained as the limit when  $k \rightarrow \infty$  of the measure  $\mu_k^\infty$  on  $\Omega^k$  given by  $d\mu_k^\infty(x_0, \dots, x_{k-1}) = d\mu_\phi(x_0) \delta_{x_1=Tx_0} \cdots \delta_{x_{k-1}=Tx_{k-2}}$ . On  $\Omega^{\mathbb{N}}$ , let  $\mathcal{F}_p$  be the  $\sigma$ -algebra of events depending only on the coordinates  $(x_j)_{j \geq p}$  (this is a decreasing sequence of  $\sigma$ -fields). We want to write the function  $K$  as a sum of reverse martingale differences with respect to this sequence. Therefore, let  $K_p = \mathbb{E}(K | \mathcal{F}_p)$  and  $D_p = K_p - K_{p+1}$ . More precisely,

$$\begin{aligned} K_p(x_p, x_{p+1}, \dots) &= \mathbb{E}(K | \mathcal{F}_p)(x_p, x_{p+1}, \dots) \\ &= \mathbb{E}(K(X_0, \dots, X_{p-1}, x_p, \dots) | X_p = x_p) \\ &= \sum_{T^p(y)=x_p} g^{(p)}(y) K(y, \dots, T^{p-1}y, x_p, \dots). \end{aligned}$$

The function  $D_p$  is  $\mathcal{F}_p$ -measurable and  $\mathbb{E}(D_p | \mathcal{F}_{p+1}) = 0$ . Moreover

$$K - \mathbb{E}(K) = \sum_{p \geq 0} D_p. \quad (7.5.6)$$

We then apply the Azuma-Hoeffding inequality (see e.g. [53, Page 68]) which says that

$$\mathbb{E} \left( e^{\sum_{p=0}^{P-1} D_p} \right) \leq e^{\frac{1}{2} \sum_{p=0}^{P-1} \|D_p\|_\infty^2}. \quad (7.5.7)$$

Therefore, the point is to obtain a good bound on  $D_p$ . This is the claim of the following lemma.

### Lemma 7.5.2.

There exists  $C_{(7.5.2)} > 0$ , depending only on  $\phi$ , such that for any  $p \in \mathbb{N}$  one has

$$\|D_p\|_\infty \leq C_{(7.5.2)} \sum_{i=0}^p \epsilon_{p-i} \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j} + \text{Lip}_{\theta,p}(K).$$

Using this lemma and applying Young's inequality for convolutions [12, p. 316] twice we obtain

$$\begin{aligned}
& \sum_{p=0}^{P-1} \|D_p\|_\infty^2 \\
& \leq 2C_{(7.5.2)}^2 \sum_{p=0}^{P-1} \left( \sum_{i=0}^p \epsilon_{p-i} \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j} \right)^2 + 2 \sum_{p=0}^{P-1} \text{Lip}_{\theta,p}(K)^2 \\
& \leq 2C_{(7.5.2)}^2 \left( \sum_{k \geq 1} \epsilon_k \right)^2 \sum_{p=0}^{P-1} \left( \sum_{j=0}^p \text{Lip}_{\theta,j}(K) \theta^{p-j} \right)^2 + 2 \sum_{p=0}^{P-1} \text{Lip}_{\theta,p}(K)^2 \\
& \leq 2 \left( C_{(7.5.2)}^2 (1-\theta)^{-2} \left( \sum_{k \geq 1} \epsilon_k \right)^2 + 1 \right) \sum_{p=0}^P \text{Lip}_{\theta,p}(K)^2.
\end{aligned}$$

**Remark 7.3.** If  $u = (u_n)_n$  and  $v = (v_n)_n$  are sequences of reals, their convolution  $u \star v$  is given by  $(u \star v)_n = \sum_{k=0}^n u_k v_{n-k}$ . Young's inequality tells us that if  $u \in \ell^p(\mathbb{N})$ ,  $v \in \ell^q(\mathbb{N})$  and  $1 \leq p, q, r \leq \infty$  with  $r^{-1} + 1 = p^{-1} + q^{-1}$ , then

$$\|u \star v\|_r \leq \|u\|_p \|v\|_q.$$

We used it twice with  $r = 2$ ,  $p = 2$  and  $q = 1$ .

Notice that by assumption and by Theorem 7.3.1 we have  $\sum_{k \geq 1} \epsilon_k < +\infty$ . Therefore, using (7.5.7) at a fixed index  $P$  and then letting  $P$  tend to infinity, we get by the dominated convergence theorem

$$\mathbb{E} \left( e^{\sum_{p \geq 0} D_p} \right) \leq e^{\frac{1}{2} \sum_{p \geq 0} \|D_p\|_\infty^2}$$

which is, in view of (7.5.6), exactly (7.4.1) with

$$C_{(7.4.1)} = 1 + C_{(7.5.2)}^2 (1-\theta)^{-2} \left( \sum_{k \geq 1} \epsilon_k \right)^2.$$

Now we are going to prove Lemma 7.5.2 by proving that  $K_p$  is close to an integral quantity. This is the content of the following lemma which is the core of the proof.

**Lemma 7.5.3.**

There exists  $C_{(7.5.3)} > 0$ , depending only on  $\phi$ , such that, for all  $p \in \mathbb{N}$ ,

$$\begin{aligned}
& \left| K_p(x_p, \dots) - \int K(y, \dots, T^{p-1}y, x_p, \dots) d\mu_\phi(y) \right| \\
& \leq C_{(7.5.3)} \sum_{i=0}^{p-1} \epsilon_{p-i} \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j}
\end{aligned}$$

where

$$C_{(7.5.3)} = C_{(7.3.1)}(c_{(7.5.4)} + 2c_{(7.5.1)} + 1).$$

*Proof of Lemma 7.5.2.* Applying Lemma 7.5.3 yields

$$\begin{aligned} & |K_p(x_p, x_{p+1}, \dots) - K_p(x'_p, x_{p+1}, \dots)| \\ & \leq 2C_{(7.5.3)} \sum_{i=0}^{p-1} \epsilon_{p-i} \sum_{j=0}^i \text{Lip}_{\theta, j}(K) \theta^{i-j} + \text{Lip}_{\theta, p}(K). \end{aligned}$$

Averaging  $K_p(x'_p, x_{p+1}, \dots)$  over the preimages of  $x'_p$  we get exactly  $K_{p+1}(x_{p+1}, \dots)$ , hence the previous bound holds for  $|D_p|$ , proving the lemma.  $\square$

*Proof of Lemma 7.5.3.* Let us fix a point  $x_*$  in  $\Omega$  and decompose  $K_p$  as

$$\begin{aligned} K_p(x_p, \dots) &= \sum_{i=0}^{p-1} \sum_{T^p(y)=x_p} g^{(p)}(y) (K(y, \dots, T^{i-1}y, T^i y, x_*, \dots, x_*, x_p, \dots) \\ & \quad - K(y, \dots, T^{i-1}y, x_*, \dots, x_*, x_p, \dots)) \\ & \quad + K(x_*, \dots, x_*, x_p, \dots). \end{aligned}$$

For fixed  $i$ , we can group together those points  $y \in T^{-p}(x_p)$  which have the same image under  $T^i$ , splitting the sum  $\sum_{T^p(y)=x_p}$  as  $\sum_{T^{p-i}(z)=x_p} \sum_{T^i(y)=z}$ . Since the Jacobian is multiplicative, one has  $g^{(p)}(y) = g^{(i)}(y)g^{(p-i)}(z)$ . Let us define two functions  $f_i$  and  $H$  as follows:

$$\begin{aligned} f_i(z) &= \sum_{T^i y=z} g^{(i)}(y) (K(y, \dots, T^{i-1}y, T^i y, x_*, \dots, x_*, x_p, \dots) \\ & \quad - K(y, \dots, T^{i-1}y, x_*, \dots, x_*, x_p, \dots)) \\ &= \sum_{T^i y=z} g^{(i)}(y) H(y, \dots, T^i y). \end{aligned}$$

Bearing in mind (7.5.2), we obtain

$$K_p(x_p, \dots) = \sum_{i=0}^{p-1} \tilde{P}_\phi^{p-i} f_i(x_p) + K(x_*, \dots, x_*, x_p, \dots).$$

Now we want to prove that  $f_i \in \mathcal{L}_\phi$  to use (7.5.3). First observe that for any  $z \in \Omega$

$$|f_i(z)| \leq \sum_{T^i y=z} g^{(i)}(y) \text{Lip}_{\theta, i}(K) d_\theta(x_*, T^i y) \leq \text{Lip}_{\theta, i}(K)$$

since  $d_\theta(x_*, T^i y) \leq 1$  and  $\sum_{T^i y=z} g^{(i)}(y) = 1$ . Hence

$$\|f_i\|_\infty \leq \text{Lip}_{\theta, i}(K).$$

We now estimate the  $d_\phi$ -Lipschitz norm of  $f_i$ . We write

$$\begin{aligned} f_i(z) - f_i(z') &= \sum (g^{(i)}(y) - g^{(i)}(y')) H(y, \dots, T^i y) \\ & \quad + \sum g^{(i)}(y') (H(y, \dots, T^i y) - H(y', \dots, T^i y')) \end{aligned} \quad (7.5.8)$$

where  $z$  and  $z'$  are two points in the same partition element, and their respective preimages  $y, y'$  are paired according to the cylinder of length  $i$  they belong to. Using the distortion control (7.5.4) we have

$$|g^{(i)}(y) - g^{(i)}(y')| \leq c_{(7.5.4)} g^{(i)}(y) d_\phi(z, z')$$

hence the first sum in (7.5.8) is bounded in absolute value by

$$c_{(7.5.4)} \text{Lip}_{\theta,i}(K) d_\phi(z, z').$$

For the second sum, substituting successively each  $T^j y$  with  $T^j y'$ , we have

$$\begin{aligned} |H(y, \dots, T^i y) - H(y', \dots, T^i y')| &\leq 2 \sum_{j=0}^i \text{Lip}_{\theta,j}(K) d_\theta(T^j y, T^j y') \\ &\leq 2 \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j} d_\theta(z, z') \\ &\leq 2c_{(7.5.1)} \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j} d_\phi(z, z') \end{aligned}$$

where we used Lemma 7.5.1 for the third inequality.

Summing over the different preimages of  $z$ , we deduce that

$$\|f_i\|_{\mathcal{L}_\phi} \leq (c_{(7.5.4)} + 2c_{(7.5.1)} + 1) \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j}.$$

**Remark 7.4.** *In the published article, the previous inequality didn't consider the complete  $\mathcal{L}_\phi$ -norm as the sum of two terms: the sup-norm and the Lipschitz constant. We correct this mistake which did not affect our final result by majoring the sup-norm.*

Therefore we can apply (7.5.3) to get

$$\left\| \tilde{P}_\phi^{p-i} f_i - \int f_i d\mu_\phi \right\|_\infty \leq C_{(7.3.1)} (c_{(7.5.4)} + 2c_{(7.5.1)} + 1) \epsilon_{p-i} \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j}.$$

Summing those bounds, one obtains

$$\begin{aligned} \left| K_p(x_p, \dots) - \sum_{i=0}^{p-1} \int f_i d\mu_\phi - K(x_*, \dots, x_*, x_p, \dots) \right| \\ \leq C_{(7.3.1)} (c_{(7.5.4)} + 2c_{(7.5.1)} + 1) \sum_{i=0}^{p-1} \epsilon_{p-i} \sum_{j=0}^i \text{Lip}_{\theta,j}(K) \theta^{i-j}. \end{aligned}$$

Finally, when one computes the sum of the integrals of  $f_i$ , there are again cancellations, leaving only  $\int K(y, \dots, T^{p-1}y, x_p, \dots) d\mu_\phi(y)$ .  $\square$

## 7.6 Applications

### 7.6.1 Birkhoff sums

Let  $f : \Omega \rightarrow \mathbb{R}$  be a  $d_\theta$ -Lipschitz function and define

$$K(x_0, \dots, x_{n-1}) = f(x_0) + \dots + f(x_{n-1})$$

whence  $K(x, Tx, \dots, T^{n-1}x) = f(x) + f(Tx) + \dots + f(T^{n-1}x) := S_n f(x)$  is the Birkhoff sum of  $f$ . Clearly,  $\text{Lip}_{\theta,i}(K) = \text{Lip}_\theta(f)$  for all  $i = 0, \dots, n-1$ . Applying Corollary 7.4.2 we immediately get

$$\mu_\phi \left( x : \left| \frac{S_n f(x)}{n} - \int f d\mu_\phi \right| \geq u \right) \leq 2 \exp \left( -c_{(7.6.1)} n u^2 \right) \quad (7.6.1)$$

for all  $n \geq 1$  and  $u \in \mathbb{R}_+$ , where

$$c_{(7.6.1)} = \frac{1}{4C_{(7.4.1)} \text{Lip}_\theta(f)^2}.$$

This bound can be compared with the large deviation asymptotics:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_\phi \left( x : \frac{S_n f(x)}{n} \geq \int f d\mu_\phi + u \right) = -I(u + \int f d\mu_\phi)$$

where  $u \geq 0$  and  $I(u) \geq 0$  is the rate function which is (strictly) convex, such that  $I(\int f d\mu_\phi) = 0$ , and equal to  $+\infty$  outside a certain finite interval  $(\underline{u}_f, \bar{u}_f)$ . We see that it has the right behavior in  $n$ . Replacing  $u$  by  $u/\sqrt{n}$  in (7.6.1) we get

$$\mu_\phi \left( x : \left| S_n f(x) - n \int f d\mu_\phi \right| \geq u\sqrt{n} \right) \leq 2 \exp \left( -c_{(7.6.1)} u^2 \right)$$

for all  $n$  and  $u > 0$ . This can be compared with the central limit theorem which holds for  $\mu_\phi$ :

$$\lim_{n \rightarrow \infty} \mu_\phi \left( x : \frac{S_n f(x) - n \int f d\mu_\phi}{\sqrt{n}} \leq +u \right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{\xi^2}{2\sigma^2}} d\xi$$

for all  $u \in \mathbb{R}$ , where  $\sigma^2 = \sigma_f^2$  is the variance of the process  $\{f(T^n x)\}_{n \geq 0}$  where  $x$  is distributed according to  $\mu_\phi$ . We can see that the previous bound is consistent with that theorem. Note that the central limit theorem is about convergence in law, whereas here we obtain a (non-asymptotic) bound from which one cannot deduce a convergence in law.

## 7.6.2 Empirical frequency of blocks

Take  $f(x) = \mathbb{1}_{[a^{0,k-1}]}(x)$  where

$$[a^{0,k-1}] = \{x \in \Omega : x^i = a^i, i = 0, \dots, k-1\}$$

is a given  $k$ -cylinder. Let

$$f_n(x, a^{0,k-1}) = \frac{\sum_{k=0}^{n-k} \mathbb{1}_{[a^{0,k-1}]}(T^k x)}{n - k + 1}.$$

This is the ‘‘empirical frequency’’ of the block  $a^{0,k-1} \in S^k$  in the orbit of  $x$  up to time  $n - k$ . By Birkhoff’s ergodic theorem, we know that, for each  $a^{0,k-1}$ ,  $f_n(x, a^{0,k-1})$  goes to  $\mu_\phi([a^{0,k-1}])$  for  $\mu_\phi$ -almost all  $x$ . The next theorem quantifies this asymptotic statement. Notice that we can control the fluctuations of  $f_n(x, a^{0,k-1})$  around  $\mu_\phi([a^{0,k-1}])$  uniformly in  $a^{0,k-1}$ .

### Theorem 7.6.1.

For all  $n \in \mathbb{N}$ , for all  $1 \leq k \leq n$  and for all  $u > 0$  we have

$$\mu_\phi \left( x : \max_{a^{0,k-1}} \left| f_n(x, a^{0,k-1}) - \mu_\phi([a^{0,k-1}]) \right| \geq \frac{(u + c\sqrt{k})\theta^{-k}}{\sqrt{n - k + 1}} \right) \leq e^{-\frac{u^2}{4C_{(7.4.1)}}}$$

where  $c = 2\sqrt{2C_{(7.4.1)} \log |S|}$ . Moreover, if  $k = k(n) = \zeta \log n$  for some

$\zeta > 0$ , then

$$\begin{aligned} \mu_\phi \left( x : \max_{a^{0,k(n)-1}} \left| f_n(x, a^{0,k(n)-1}) - \mu_\phi([a^{0,k(n)-1}]) \right| \geq \frac{(u + c' \sqrt{\log n}) n^{\zeta |\log \theta|}}{\sqrt{n - k(n) + 1}} \right) \\ \leq e^{-\frac{u^2}{4C_{(7.4.1)}}} \end{aligned}$$

where  $c' = 2\sqrt{2\zeta C_{(7.4.1)} \log |S|}$ .

*Proof.* Define the function  $K : \Omega^{n-k+1} \rightarrow \mathbb{R}$  by

$$K(x_0, \dots, x_{n-k}) = \max_{a^{0,k-1}} Z(a^{0,k-1}; x_0, \dots, x_{n-k})$$

where

$$Z(a^{0,k-1}; x_0, \dots, x_{n-k}) = \left| \frac{\sum_{j=0}^{n-k} \mathbf{1}_{[a^{0,k-1}]}(x_j)}{n - k + 1} - \mu_\phi([a^{0,k-1}]) \right|.$$

It is left to the reader to check that  $\text{Lip}_{\theta,j}(K) = \frac{\text{Lip}_\theta(f)}{n-k+1} = \frac{1}{\theta^k(n-k+1)}$ , so we get immediately from 7.6.1

$$\begin{aligned} \mu_\phi \left( x \in \Omega : K(x, Tx, \dots, T^{n-k}x) \geq u + \int K(y, Ty, \dots, T^{n-k}y) d\mu_\phi(y) \right) \\ \leq \exp \left( -\frac{\theta^{2k}}{4C_{(7.4.1)}} (n - k + 1) u^2 \right) \end{aligned}$$

for all  $n \geq 1$  and  $u > 0$ . To complete the proof, we need a good upper bound for  $\int K(y, Ty, \dots, T^{n-k}y) d\mu_\phi(y)$ . Actually, this can be done by using again the Gaussian concentration bound. Using (7.4.1) and Jensen's inequality we get for any  $\xi > 0$

$$\begin{aligned} \exp \left( \xi \int K(x, Tx, \dots, T^{n-k}x) d\mu_\phi(x) \right) \\ \leq \int \exp \left( \xi \max_{a^{0,k-1}} Z(a^{0,k-1}; x, Tx, \dots, T^{n-k}x) \right) d\mu_\phi(x) \\ \leq \sum_{a^{0,k-1} \in S^k} \int \exp \left( \xi Z(a^{0,k-1}; x, Tx, \dots, T^{n-k}x) \right) d\mu_\phi(x) \\ \leq 2|S|^k \exp \left( \frac{C_{(7.4.1)} \theta^{-2k} \xi^2}{n - k + 1} \right). \end{aligned}$$

The third inequality is obtained by using the trivial inequality

$$e^{\max_{i=1}^p a_i} \leq \sum_{i=1}^p e^{a_i}.$$

Taking logarithms on both sides and then dividing by  $\xi$ , we have

$$\int K(x, Tx, \dots, T^{n-k}x) d\mu_\phi(x) \leq \frac{\log 2 + k \log |S|}{\xi} + \frac{C_{(7.4.1)} \theta^{-2k} \xi}{n - k + 1}.$$

There is a unique  $\xi > 0$  minimizing the right-hand side, hence

$$\int K(x, Tx, \dots, T^{n-k}x) d\mu_\phi(x) \leq 2\theta^{-k} \sqrt{\frac{C_{(7.4.1)}(k+1) \log |S|}{n - k + 1}}$$

where we used that  $\log 2 \leq \log |S|$ . Hence we get the desired estimate.  $\square$

Note that  $\log |S|$  is the topological entropy of the full shift with alphabet  $S$ .

### 7.6.3 Hitting times and entropy

For  $x, y \in \Omega$ , let

$$T_{x^0, n-1}(y) = \inf\{j \geq 1 : y^{j, j+n-1} = x^{0, n-1}\}.$$

This is the first time that the  $n$  first symbols of  $x$  appear in  $y$ . We assume that  $\phi$  satisfies

$$\text{var}_n(\phi) = O\left(\frac{1}{n^\alpha}\right) \quad \text{for some } \alpha > 2. \quad (7.6.2)$$

One can prove (see [22]) that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log T_{x^0, n-1}(y) = h(\mu_\phi), \quad \text{for } \mu_\phi \otimes \mu_\phi\text{-almost every } (x, y).$$

Roughly, this means that, if we pick  $x$  and  $y$  independently, each one according to  $\mu_\phi$ , then the time it takes to see the first  $n$  symbols of  $x$  appearing in  $y$  for the first time is  $\approx e^{nh(\mu_\phi)}$ .

#### Theorem 7.6.2.

If  $\phi$  satisfies (7.6.2), then there exist strictly positive constants  $c_1, c_2$  and  $u_0$  such that, for all  $n$  and for all  $u > u_0$ ,

$$(\mu_\phi \otimes \mu_\phi)\left\{(x, y) : \frac{1}{n} \log T_{x^0, n-1}(y) \geq h(\mu_\phi) + u\right\} \leq c_1 e^{-c_2 n u^2}$$

and

$$(\mu_\phi \otimes \mu_\phi)\left\{(x, y) : \frac{1}{n} \log T_{x^0, n-1}(y) \leq h(\mu_\phi) - u\right\} \leq c_1 e^{-c_2 n u}.$$

These bounds were obtained in [18] when  $\phi$  is Lipschitz. Observe that the probability of being above  $h(\mu_\phi)$  is bounded above by  $c_1 e^{-c_2 n u^2}$ , whereas the probability of being below  $h(\mu_\phi)$  is bounded above by  $c_1 e^{-c_2 n u}$ . The proof of this theorem being very similar to that given in [18], we omit the details and only sketch it. We cannot directly deal with  $T_{x^0, n-1}(y)$  but we have  $\log T_{x^0, n-1}(y) = \log(T_{x^0, n-1}(y)\mu_\phi([x^{0, n-1}])) - \log \mu_\phi([x^{0, n-1}])$ . Then we use Theorem 7.4.2 for  $\psi = -\phi$ , assuming (without loss of generality) that  $P(\phi) = 0$ , that is,  $h(\mu_\phi) = -\int \phi d\mu_\phi$ , because we can control uniformly in  $x$  the approximation  $-\log \mu_\phi([x^{0, n-1}]) \approx S_n(-\phi)(x)$ . To control the other term, we use that the law of  $T_{x^0, n-1}(y)\mu_\phi([x^{0, n-1}])$  is well approximated by an exponential law.

Another estimator of  $h(\mu_\phi)$  is the so-called plug-in estimator. We could also obtain concentration bounds for it in the spirit of [18].

### 7.6.4 Speed of convergence of the empirical measure

Instead of looking at the frequency of a block  $a_1^k$  we can consider a global object, namely the empirical measure

$$\mathcal{E}_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x}.$$

For  $\mu_\phi$ -almost every  $x$ , we know that

$$\frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x} \xrightarrow{n \rightarrow \infty} \mu_\phi$$



where the convergence is in the weak topology on the space of probability measures  $\mathcal{M}(\Omega)$  on  $\Omega$ . This is a consequence of Birkhoff's ergodic theorem. The natural question is: how fast does this convergence takes place? We can answer this question by using the Kantorovich distance  $d_K$  which metrizes weak topology on  $\mathcal{M}(\Omega)$ :

$$d_K(\nu_1, \nu_2) = \sup \left\{ \int f d\nu_1 - \int f d\nu_2 : f : \Omega \rightarrow \mathbb{R} \text{ such that } \text{Lip}_\theta(f) = 1 \right\}.$$

We have the following result.

**Theorem 7.6.3.**

For all  $u > 0$  and all  $n \geq 1$  we have

$$\mu_\phi \left( x : \left| d_K(\mathcal{E}_n(x), \mu_\phi) - \int d_K(\mathcal{E}_n(y), \mu_\phi) d\mu_\phi(y) \right| \geq u \right) \leq 2 e^{-c_{(7.6.3)} n u^2} \quad (7.6.3)$$

where  $c_{(7.6.3)} = (4C_{(7.4.1)})^{-1}$ .

*Proof.* Let

$$K(x_0, \dots, x_{n-1}) = \sup \left\{ \frac{1}{n} \sum_{j=0}^{n-1} f(x_j) - \int f d\mu_\phi : f : \Omega \rightarrow \mathbb{R} \text{ with } \text{Lip}_\theta(f) = 1 \right\}.$$

Of course,  $K(x, Tx, \dots, T^{n-1}x) = d_K(\mathcal{E}_n(x), \mu_\phi)$ . It is left to the reader to check that

$$\text{Lip}_{\theta,i}(K) \leq \frac{1}{n}, \quad i = 0, \dots, n-1.$$

The result follows at once by applying inequality (7.4.3).  $\square$

It is natural to ask for a good upper bound for  $\int d_K(\mathcal{E}_n(y), \mu_\phi) d\mu_\phi(y)$  because this would give a control on the fluctuations of  $d_K(\mathcal{E}_n(x), \mu_\phi)$  around 0. Getting such a bound turns out to be difficult. In [14, Section 8] it is proved that

$$\int d_K(\mathcal{E}_n(y), \mu_\phi) d\mu_\phi(y) \leq \frac{1}{n^{2(1+\log|A|)}}.$$

For two positive sequences  $(a_n), (b_n)$ ,  $a_n \preceq b_n$  means that  $\limsup_n \frac{\log a_n}{\log b_n} \leq 1$ . One could in principle get a non-asymptotic but messy bound.

### 7.6.5 Relative entropy, $\bar{d}$ -distance and speed of Markov approximation

Given  $n \in \mathbb{N}$  and  $x^{0,n-1}, y^{0,n-1} \in A^n$  the (non-normalized) Hamming distance between  $x$  and  $y$  is

$$\bar{d}_n(x^{0,n-1}, y^{0,n-1}) = \sum_{i=0}^{n-1} \mathbb{1}_{\{x^i \neq y^i\}}. \quad (7.6.4)$$

Now, given two shift-invariant probability measures  $\mu, \nu$  on  $\Omega$ , denote by  $\mu_n$  and  $\nu_n$  their projections on  $A^n$ , and define their  $\bar{d}_n$ -distance by

$$\bar{d}_n(\mu_n, \nu_n) = \inf \sum_{x^{0,n-1} \in A^n} \sum_{y^{0,n-1} \in A^n} \bar{d}_n(x, y) \mathbb{P}_n(x^{0,n-1}, y^{0,n-1})$$

where the infimum is taken over all the joint shift-invariant probability distributions  $\mathbb{P}_n$  on  $A^n \times A^n$  such that  $\sum_{y^{0,n-1} \in A^n} \mathbb{P}_n(x^{0,n-1}, y^{0,n-1}) = \mu_n(x^{0,n-1})$  and  $\sum_{x^{0,n-1} \in A^n} \mathbb{P}_n(x^{0,n-1}, y^{0,n-1}) = \nu_n(y^{0,n-1})$ . By [78, Theorem I.9.6, p. 92], the following limit exists:

$$\bar{d}(\mu, \nu) := \lim_{n \rightarrow \infty} \frac{1}{n} \bar{d}_n(\mu_n, \nu_n) \quad (7.6.5)$$

and defines a distance on the set of shift-invariant probability measures. It induces a finer topology than the weak topology and, in particular, the  $\bar{d}$ -limit of ergodic measures is ergodic, and the entropy is  $\bar{d}$ -continuous on the class of ergodic measures.<sup>1</sup>

Next, given  $n \in \mathbb{N}$  and a shift-invariant probability measure  $\nu$  on  $\Omega$ , define the  $n$ -block relative entropy of  $\nu$  with respect to  $\mu_\phi$  by

$$H_n(\nu | \mu_\phi) = \sum_{x^{0,n-1} \in A^n} \nu_n(x^{0,n-1}) \log \frac{\nu_n(x^{0,n-1})}{\mu_{\phi,n}(x^{0,n-1})}.$$

One can easily prove that the following limit exists and defines the relative entropy of  $\nu$  with respect to  $\mu_\phi$ :

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_n(\nu_n | \mu_{\phi,n}) =: h(\nu | \mu_\phi) = P(\phi) - \int \phi d\nu - h(\nu) \quad (7.6.6)$$

where  $P(\phi)$  is the topological pressure of  $\phi$ :

$$P(\phi) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{a^{0,n-1} \in A^n} e^{\sup \{S_n \phi(x) : x \in [a^{0,n-1}]\}}.$$

This limit exists for any continuous  $\phi$ . (To prove (7.6.6), we use that there exists a positive sequence  $(\varepsilon_n)_n$  going to 0 such that, for any  $a^{0,n-1} \in A^n$  and any  $x \in [a^{0,n-1}]$ ,  $\mu_\phi([a^{0,n-1}]) / \exp(-nP(\phi) + S_n \phi(x))$  is bounded below by  $\exp(-n\varepsilon_n)$  and above by  $\exp(-n\varepsilon_n)$ .) By the variational principle,  $h(\nu | \mu_\phi) \geq 0$  with equality if and only if  $\nu = \mu_\phi$  (recall that  $\mu_\phi$  is the unique equilibrium state of  $\phi$ ). We refer to [83] for details. We can now formulate the first theorem of this section.

**Theorem 7.6.4.**

For every shift-invariant probability measure  $\nu$  on  $\Omega$  and for all  $n \in \mathbb{N}$ , we have

$$\bar{d}_n(\nu_n, \mu_{\phi,n}) \leq c_{(7.6.4)} \sqrt{n H_n(\nu_n | \mu_{\phi,n})} \quad (7.6.7)$$

where  $c_{(7.6.4)} = \sqrt{2C_{(7.4.1)}}$ . In particular

$$\bar{d}(\nu, \mu_\phi) \leq c_{(7.6.4)} \sqrt{h(\nu | \mu_\phi)}. \quad (7.6.8)$$

*Proof.* For a function  $f : S^n \rightarrow \mathbb{R}$ , define for each  $i = 0, \dots, n-1$

$$\delta_i(f) = \sup\{|f(a^{0,n-1}) - f(b^{0,n-1})| : a^j = b^j, \forall j \neq i\}.$$

We obviously have that for all  $a^{0,n-1}, b^{0,n-1} \in S^n$

$$|f(a^{0,n-1}) - f(b^{0,n-1})| \leq \sum_{j=0}^{n-1} \mathbb{1}_{\{a^j \neq b^j\}} \delta_j(f).$$

<sup>1</sup>These two properties are false in the weak topology.

A function  $f : S^n \rightarrow \mathbb{R}$  such that  $\delta_j(f) = 1$ ,  $j = 0, \dots, n-1$  is 1-Lipschitz for the Hamming distance (7.6.4). We now consider the set of functions

$$\mathcal{H}(n, \phi) = \left\{ f : S^n \rightarrow \mathbb{R} : f \text{ 1-Lipschitz for } \bar{d}_n, \int_{A^n} f \, d\mu_{\phi, n} = 0 \right\}.$$

We can identify a function  $f \in \mathcal{H}(n, \phi)$  with a function  $\tilde{f} : \Omega^n \rightarrow \mathbb{R}$  in a natural way:  $\tilde{f}(x_0, \dots, x_{n-1}) = f(\pi(x_0), \dots, \pi(x_{n-1}))$  where  $\pi : \Omega \rightarrow S$  is defined by  $\pi(x) = x^0$ . We obviously have  $\int \tilde{f} \, d\mu_\phi = 0$  and it is easy to check that  $\text{Lip}_j(\tilde{f}) = \delta_j(f) = 1$ ,  $j = 0, \dots, n-1$ . Therefore we can apply the Gaussian concentration bound (7.4.1) to get

$$\int_{A^n} e^{\xi f} \, d\mu_{\phi, n} \leq e^{C(7.4.1)n\xi^2}, \text{ for all } f \in \mathcal{H}(n, \phi) \text{ and for all } \xi \in \mathbb{R}. \quad (7.6.9)$$

We now apply an abstract result [8, Theorem 3.1] which says that (7.6.9) is equivalent to

$$\bar{d}(\nu_n, \mu_{\phi, n}) \leq \sqrt{2C(7.4.1)nH_n(\nu_n|\mu_{\phi, n})} \text{ for all probability measures } \nu_n \text{ on } S^n.$$

Hence (7.6.7) is proved. To get (7.6.8), divide by  $n$  on both sides and take the limit  $n \rightarrow \infty$  and use (7.6.5) and (7.6.6).  $\square$

We now give an application of inequality (7.6.8). Let

$$\phi_1(x) = \log \mu_\phi(x^0) \quad \text{and} \quad \phi_n(x) = \log \mu_\phi(x^{n-1}|x^{0, n-2}), \quad n \geq 2.$$

The equilibrium state for  $\phi_n$  is a  $(n-1)$ -step Markov measure. One can prove that in the weak topology  $(\mu_{\phi_n})_n$  converges to  $\mu_\phi$ , but one cannot get any speed of convergence. We get the following upper bound on the speed of convergence of  $(\mu_{\phi_n})_n$  to  $\mu_\phi$  in the finer  $\bar{d}$  topology.

**Corollary 7.6.1.**

Assume, without loss of generality, that  $\phi$  is normalized in the sense that

$$\sum_{a \in S} e^{\phi(ax)} = 1, \quad \forall x \in \Omega.$$

Then there exists  $n_\phi \geq 1$  such that, for all  $n \geq n_\phi$ , we have

$$\bar{d}(\mu_{\phi_n}, \mu_\phi) \leq \rho_\phi \text{var}_n(\phi) \quad (7.6.10)$$

where

$$\rho_\phi = \sqrt{2|S|C(7.4.1)}(e-1)e^{\frac{3}{2}\|\phi\|_\infty}.$$

We explained how to normalize a potential in Subsection 7.5.1.

*Proof.* Using (7.6.6) and the variational principle we get

$$h(\mu_{\phi_n}|\mu_\phi) = - \int \phi \, d\mu_{\phi_n} - h(\mu_{\phi_n}) = \int (\phi_n - \phi) \, d\mu_{\phi_n}. \quad (7.6.11)$$

Indeed, since  $\phi$  and  $\phi_n$  are normalized, we have in particular that  $P(\phi) = P(\phi_n) = 0$ , and by the variational principle  $h(\mu_{\phi_n}) = - \int \phi_n \, d\mu_{\phi_n}$ . Now

$$\begin{aligned} \int (\phi_n - \phi) \, d\mu_{\phi_n} &= \int \log \left( \frac{e^{\phi_n}}{e^\phi} \right) \, d\mu_{\phi_n} = \int \log \left( 1 + \frac{e^{\phi_n} - e^\phi}{e^\phi} \right) \, d\mu_{\phi_n} \\ &\leq \int \frac{e^{\phi_n} - e^\phi}{e^\phi} \, d\mu_{\phi_n} \end{aligned} \quad (7.6.12)$$

where we used the inequality  $\log(1 + u) \leq u$  for all  $u > -1$ . Now using the shift-invariance of  $\mu_{\phi_n}$  and replacing  $e^{\phi_n}$  by  $e^{\phi_n} - e^\phi + e^\phi$  we get

$$\begin{aligned} \int \frac{e^{\phi_n} - e^\phi}{e^\phi} d\mu_{\phi_n} &= \int d\mu_{\phi_n}(x) \sum_{a \in S} e^{\phi_n(ax)} \frac{e^{\phi_n(ax) - e^\phi(ax)}}{e^{\phi(ax)}} \\ &= \int d\mu_{\phi_n}(x) \sum_{a \in S} \frac{(e^{\phi_n(ax)} - e^{\phi(ax)})^2}{e^{\phi(ax)}} + \int d\mu_{\phi_n}(x) \sum_{a \in S} (e^{\phi_n(ax)} - e^{\phi(ax)}) \\ &\leq |S| e^{\|\phi\|_\infty} (\|e^{\phi_n} - e^\phi\|_\infty)^2 \end{aligned} \quad (7.6.13)$$

where we used that  $\sum_{a \in S} (e^{\phi_n(ax)} - e^{\phi(ax)}) = 0$ . Combining (7.6.8), (7.6.11), (7.6.12) and (7.6.13) we thus obtain

$$\bar{d}(\mu_{\phi_n}, \mu_\phi) \leq \sqrt{2|S| C_{(7.4.1)} e^{\|\phi\|_\infty}} \|e^{\phi_n} - e^\phi\|_\infty.$$

It remains to estimate  $\|e^{\phi_n} - e^\phi\|_\infty$  in terms of  $\text{var}_n(\phi)$ . We have

$$\|e^{\phi_n} - e^\phi\|_\infty = \|e^{\phi_n} - e^\phi\|_\infty \leq e^{\|\phi\|_\infty} \|e^{\phi - \phi_n} - 1\|_\infty \leq (e - 1) e^{\|\phi\|_\infty} \|\phi - \phi_n\|_\infty$$

provided that  $\|\phi - \phi_n\|_\infty < 1$ , where we used the inequality  $|e^u - 1| \leq (e - 1)|u|$  valid for  $|u| < 1$ . Finally, since  $\|\phi - \phi_n\|_\infty \leq \text{var}_n(\phi)$ , we define  $n_\phi$  to be the smallest integer such that  $\text{var}_n(\phi) < 1$  and we can take

$$\rho_\phi = \sqrt{2|S| C_{(7.4.1)} (e - 1) e^{\frac{3}{2}\|\phi\|_\infty}}.$$

We thus proved (7.6.10).  $\square$

Let us mention the paper [35] in which the authors obtain the same bound for the speed of convergence of Markov approximations, up to the constant. Their approach is a direct estimation of  $\bar{d}(\mu_{\phi_n}, \mu_\phi)$  by using a coupling method. The point here is to obtain the same speed of convergence as an easy corollary of inequality (7.6.8). Let us remark that from (7.5.8) we get a worse result since we end up with a bound proportional to  $\sqrt{\text{var}_n(\phi)}$ . The trick which leads to the correct bound was told us by Daniel Takahashi.

## 7.6.6 Shadowing of orbits

Let  $A$  be a Borel subset of  $\Omega$  such that  $\mu_\phi(A) > 0$  and define for all  $n \in \mathbb{N}$

$$\mathcal{S}_A(x, n) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} d_\theta(T^j x, T^j y)$$

A basic example of such a set  $A$  is a cylinder set  $[a^{0, k-1}]$ . The quantity  $\mathcal{S}_A(x, n)$ , which lies between 0 and 1, measures how we can trace, in the best possible way, the orbit of some initial condition not in  $A$  by an orbit starting in  $A$ .

### Theorem 7.6.5.

For any Borel subset  $A \subset \Omega$  such that  $\mu_\phi(A) > 0$ , for any  $n \in \mathbb{N}$  and for any  $u > 0$

$$\mu_\phi \left\{ x \in \Omega : \mathcal{S}_A(x, n) \geq \frac{u_A + u}{\sqrt{n}} \right\} \leq e^{-\frac{u^2}{4C_{(7.4.1)}}}$$

where

$$u_A = 2\sqrt{-C_{(7.4.1)} \ln \mu_\phi(A)}.$$

We give a shorter and simpler proof than in [24].

*Proof.* Let  $K(x_0, \dots, x_{n-1}) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} d_\theta(x_j, T^j y)$ . One can easily check that

$$\text{Lip}_{\theta,i}(K) = \frac{1}{n}, \quad \forall i = 0, \dots, n-1.$$

It follows from (7.4.2) that

$$\mu_\phi \left\{ \mathcal{S}_A(x, n) \geq \int \mathcal{S}_A(y, n) d\mu_\phi(y) + \frac{u}{\sqrt{n}} \right\} \leq e^{-\frac{u^2}{4C_{(7.4.1)}}}$$

for all  $n \geq 1$  and for all  $u > 0$ . We now need an upper bound for  $\int \mathcal{S}_A(y, n) d\mu_\phi(y)$ . We simply observe that by (7.4.1) and the definition of  $\mathcal{S}_A(\cdot, n)$

$$\begin{aligned} \mu_\phi(A) &= \int e^{-\xi \mathcal{S}_A(x, n)} \mathbb{1}_A(x) d\mu_\phi(x) \leq \int e^{-\xi \mathcal{S}_A(x, n)} d\mu_\phi(x) \\ &\leq e^{-\xi \int \mathcal{S}_A(y, n) d\mu_\phi(y)} e^{\frac{C_{(7.4.1)} \xi^2}{n}} \end{aligned}$$

for all  $\xi > 0$ . Hence

$$\int \mathcal{S}_A(y, n) d\mu_\phi(y) \leq \frac{C_{(7.4.1)} \xi}{n} + \frac{1}{\xi} \ln(\mu_\phi(A)^{-1})$$

Optimizing this bound over  $\xi > 0$  gives

$$\int \mathcal{S}_A(y, n) d\mu_\phi(y) \leq 2\sqrt{\frac{C_{(7.4.1)} \ln(\mu_\phi(A)^{-1})}{n}}.$$

The theorem follows at once.  $\square$

### 7.6.7 Almost-sure central limit theorem

It was proved in [66, Chapter 2] that  $(\Omega, T, \mu_\phi)$  satisfies the central limit theorem for the class of  $d_\theta$ -Lipschitz functions  $f : \Omega \rightarrow \mathbb{R}$  such  $\int f d\mu_\phi = 0$ , that is, for any such  $f$  the process  $\{f \circ T^n\}_{n \geq 0}$  satisfies

$$\mu_\phi \left( x : \frac{S_k f(x)}{\sqrt{k}} \leq u \right) = \int \mathbb{1}_{\left\{ \frac{S_k f(x)}{\sqrt{k}} \leq t \right\}} d\mu_\phi(x) \xrightarrow{k \rightarrow \infty} G_{0, \sigma^2(f)}((-\infty, t]) \quad (7.6.14)$$

where

$$\sigma^2(f) = \int f^2 d\mu_\phi + 2 \sum_{i \geq 1} \int f \cdot f \circ T^i d\mu_\phi \in [0, +\infty).$$

If  $\sigma^2(f) > 0$ ,  $G_{0, \sigma^2}$  denotes the law of a Gaussian random variable with mean 0 and variance  $\sigma^2(f)$ , that is,

$$dG_{0, \sigma^2(f)}(u) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2(f)}} du, \quad u \in \mathbb{R}.$$

When  $\sigma^2(f) = 0$  we set  $G_{0,0} = \delta_0$ , the Dirac mass at zero.

**Remark 7.5.** In fact, a more general statement was proved in [66, Chapter 1.2]. Namely, (7.6.14) holds when  $\phi$  is such that  $\sum_k \epsilon_k < +\infty$  and  $f \in \mathcal{L}_\phi$ .

Now, for each  $N \geq 1$  and  $x \in \Omega$ , define the probability measure

$$\mathcal{A}_N(x) = \frac{1}{L_N} \sum_{n=1}^N \frac{1}{n} \delta_{\frac{S_n f(x)}{\sqrt{n}}} \quad (7.6.15)$$

where  $L_N = \sum_{n=1}^N \frac{1}{n}$  and where, as usual,  $\delta_u$  is the Dirac mass at point  $u \in \mathbb{R}$ . Of course,  $L_N = \log N + \mathcal{O}(1)$ . Notice that  $\mathcal{A}_N$  is a random probability measure. Finally, the Wasserstein distance between two probability measures  $\nu, \nu'$  on the Borel sigma-algebra  $\mathcal{B}(\mathbb{R})$  is

$$W_1(\nu, \nu') = \inf_{\pi \in \Pi(\nu, \nu')} \int d_\theta(x, x') d\pi(x, x')$$

where the infimum is taken over all probability measures such that

$$\int \pi(B, x') dx' = \nu(B) \quad \text{and} \quad \int \pi(x, B) dx = \nu'(B)$$

for any Borel subset of  $\mathbb{R}$ . By the Kantorovich-Rubinstein duality theorem,  $W_1(\nu, \nu')$  is equal to the Kantorovich distance which is the supremum of  $\int \ell d\nu - \int \ell d\nu'$  over the set of 1-Lipschitz functions  $\ell : \mathbb{R} \rightarrow \mathbb{R}$ . We refer to [32] for background and proofs.

Now we can formulate the almost-sure central limit theorem.

**Theorem 7.6.6.**

Let  $f : \Omega \rightarrow \mathbb{R}$  be a  $d_\theta$ -Lipschitz function. Then, for  $\mu_\phi$  almost every  $x \in \Omega$ , we have

$$W_1(\mathcal{A}_N(x), G_{0, \sigma^2(f)}) \xrightarrow{N \rightarrow +\infty} 0.$$

We make several comments. Recall that the Wasserstein distance metrizes the weak topology on the set of probability measures  $\nu$  on  $\mathcal{B}(\mathbb{R})$ . Moreover, if  $(\nu_n)_{n \geq 1}$  is a sequence of probability measures on  $\mathcal{B}(\mathbb{R})$  and  $\nu$  a probability measure on  $\mathcal{B}(\mathbb{R})$ , then

$$\lim_{n \rightarrow \infty} W_1(\nu_n, \nu) = 0 \iff \nu_n \xrightarrow{\text{law}} \nu \quad \text{and} \quad \int |u| d\nu_n(u) \xrightarrow{n \rightarrow +\infty} \int |u| d\nu(u)$$

where “ $\xrightarrow{\text{law}}$ ” means weak convergence of probability measures on  $\mathcal{B}(\mathbb{R})$ .

To compare with (7.6.14), observe that Theorem 7.6.6 implies that for  $\mu_\phi$ -almost every  $x$ ,  $\mathcal{A}_N(x) \xrightarrow{\text{law}} G_{0, \sigma^2(f)}$ , which in turn implies that

$$\int \mathbb{1}_{\{u \leq t\}} d\mathcal{A}_N(u) = \frac{1}{L_N} \sum_{n=1}^N \frac{1}{n} \mathbb{1}_{\{S_n f / \sqrt{n} \leq t\}} \xrightarrow{N \rightarrow +\infty} G_{0, \sigma^2(f)}((-\infty, t]).$$

Therefore, the expectation with respect to  $\mu_\phi$  in (7.6.14) is replaced by a pathwise logarithmic average in the almost-sure central limit theorem.

*Proof.* The proof follows from an abstract theorem proved in [15]. In words, that theorem says the following. Let  $(X_n)_{n \geq 0}$  be a stochastic stationary process where the  $X_n$ 's are random variables taking values in  $\Omega$ . Assume that if  $f : \Omega \rightarrow \mathbb{R}$  is  $d$ -Lipschitz and such that  $\mathbb{E}[f(X_0)] = 0$ , then it satisfies the central limit theorem, that is, for all  $u \in \mathbb{R}$ ,

$$\mathbb{P} \left( \frac{\sum_{j=0}^{n-1} f(X_j)}{\sqrt{n}} \leq u \right) \xrightarrow{n \rightarrow \infty} G_{0, \sigma^2(f)}((-\infty, u])$$

where  $\sigma^2(f) := \mathbb{E}[f^2(X_0)] + 2 \sum_{\ell \geq 1} \mathbb{E}[f(X_0)f(X_\ell)]$  is assumed to be  $\neq 0$ . Moreover, assume that the process  $(X_n)_{n \geq 0}$  satisfies the following variance inequality: There exists  $C > 0$  such that for all separately  $d$ -Lipschitz functions  $K : \Omega^n \rightarrow \mathbb{R}$  for some distance  $d$  on  $\Omega$ ,

$$\mathbb{E}[(K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})])^2] \leq C \sum_{i=0}^{n-1} \text{Lip}_i(K)^2.$$

Then, the conclusion is that, almost surely,

$$\frac{1}{L_N} \sum_{n=1}^N \frac{1}{n} \delta_{\frac{X_0 + \dots + X_{n-1}}{\sqrt{n}}}$$

converges in Wasserstein distance (or, equivalently, in Kantorovich distance) to  $G_{0, \sigma^2(f)}((-\infty, u])$ . We apply this abstract theorem to the process  $(x, Tx, \dots)$  where  $x \in \Omega$  is distributed according to  $\mu_\phi$  with  $\Omega = S^{\mathbb{N}}$  and  $d = d_\theta$ . Since we have (7.6.14) and (7.4.4), the theorem follows.  $\square$

**Remark 7.6.** *The previous result relies only upon the variance inequality (7.4.4), which is much weaker than the Gaussian concentration bound of Theorem 7.4.1. On the one hand, the variance inequality (7.4.4) should be true for less regular potentials than the ones we consider here. On the other hand, the Gaussian concentration bound should provide a strengthening of Theorem 7.6.6, namely a speed of convergence.*

## 7.7 Appendix

### 7.7.1 Cones and projective metrics

Here, we recall the definitions and properties of the powerful method initiated by Birkhoff to study linear operators and we give the detailed proof of V. Maume's thesis written in french ([66]). (In her thesis, she estimated more precisely the speed of convergence of the transfer operator to the equilibrium state than in [47].) Like the classical coupling approach, this strategy was widely applied to estimate decays of correlation of dynamical systems. She constructed a sequence of cones of Lipschitz functions in which she obtained a non-uniform contraction of the Ruelle-Perron-Frobenius operator.

**Definition 7.7.1.** *[Cone set]*

Let  $B$  be a vector space and  $C \subset B$  a convex cone, i.e.,

1. Let  $x \in C$  then  $\forall \lambda > 0, \lambda x \in C$ ,
2.  $C$  is a convex set,
3.  $C \cap -C = \emptyset$ ,
4. For all sequences of real numbers  $(\alpha_n)_{n \geq 0}$  converging to  $\alpha$  and if  $\forall n \in \mathbb{N}, x - \alpha_n y \in C$  then  $x - \alpha y \in C$ .

We define the pseudo-metric  $\theta$  associated to this cone this way

**Definition 7.7.2.**

Let  $(x, y) \in C$  and

$$\begin{aligned}\mu(x, y) &= \sup\{\beta > 0, \beta x - y \in C\}, \\ \lambda(x, y) &= \inf\{\alpha > 0, y - \alpha x \in C\}.\end{aligned}$$

Assuming that  $\mu(x, y) = \infty$  and  $\lambda(x, y) = 0$  if the corresponding sets are empty. We define the pseudo-metric  $\theta_C$  in the following way

$$\theta_C(x, y) = \log \frac{\mu(x, y)}{\lambda(x, y)}. \quad (7.7.1)$$

It's a pseudo-metric because  $\theta(x, y)$  is not necessarily finite. Let  $C, C'$  be two cones and  $P : C \rightarrow C'$  be a linear operator such that  $PC \subset C'$ . We denote by  $\Delta$  the diameter of  $PC$  into  $C'$  as follows

$$\Delta := \sup_{f, g \in C} \theta_{C'}(Pf, Pg).$$

**Proposition 7.7.1.**

For all functions  $f, g \in C$  we have

$$\theta_{C'}(Pf, Pg) \leq \tanh\left(\frac{\Delta}{4}\right) \theta_C(f, g). \quad (7.7.2)$$

*Proof.* Let  $f, g \in C$ . First, if  $\theta_C(f, g) = \infty$  then (7.7.2) is trivial. Let us assume that  $\theta_C(f, g) = \log \frac{\mu}{\lambda} < \infty$ , then  $\mu < \infty$  and  $\lambda \neq 0$ . By 4 and the linearity property of  $P$ , the following holds for  $\mu$  and  $\lambda$

$$\mu Pf - Pg \in C' \quad \text{and} \quad Pg - \lambda Pf \in C'.$$

Then, one has  $\theta_{C'}(Pf, Pg) \leq \theta_C(f, g)$  which implies  $\Delta = \infty$ . Now, if  $\Delta < \infty$ , then

$$\theta_{C'}(P(\mu f - g), P(g - \lambda f)) \leq \Delta.$$

By definition, there exist positive real numbers  $\alpha, \beta$  such that  $\log \frac{\beta}{\alpha} \Delta$  and

$$\begin{aligned}\beta P(\mu f - g) - P(g - \lambda f) &\in C', \\ P(g - \lambda f) - \alpha P(\mu f - g) &\in C'.\end{aligned}$$

On the other hand, one has

$$\begin{aligned}\frac{\beta\mu + \lambda}{\beta + 1} Pf - Pg &\in C', \\ Pg - \frac{\lambda + \alpha\mu}{\alpha + 1} Pf &\in C'.\end{aligned}$$



Hence, by the projective metric definition,  $Pf$  and  $Pg$  satisfy

$$\begin{aligned}
\theta_{C'}(Pf, Pg) &\leq \log \frac{(\beta\mu + \lambda)(\alpha + 1)}{(\lambda + \alpha\mu)(\beta + 1)} \\
&= \log \frac{\beta + e^{-\theta_C(f,g)}}{\alpha + e^{-\theta_C(f,g)}} - \log \frac{\beta + 1}{\alpha + 1} \\
&= \int_0^{\theta_C(f,g)} \frac{(\beta - \alpha)e^{-x}}{(e^{-x} + \alpha)(e^{-x} + \beta)} dx \\
&\leq \theta_C(f, g) \sup_{0 \leq t \leq 1} \frac{(\beta - \alpha)t}{(t + \alpha)(t + \beta)} \\
&\leq \theta_C(f, g) \frac{1 - \frac{\alpha}{\beta}}{(1 + \sqrt{\frac{\alpha}{\beta}})^2} \\
&\leq \theta_C(f, g) \tanh \frac{\Delta}{4}.
\end{aligned}$$

□

**Definition 7.7.3.**

We say that a norm  $\|\cdot\|$  on  $B$  is adapted to  $C$ , if for all  $f, g \in B$  such that  $f + g \in C$  and  $f - g \in C$  then  $\|f\| \leq \|g\|$ . We say that  $\rho$  is an homogeneous form adapted to  $C$  if  $\rho : C \rightarrow \mathbb{R}^+$  satisfies: For  $\lambda > 0$ ,  $\rho(\lambda f) = \lambda\rho(f)$  and if  $f - g \in C$  then  $\rho(f) \leq \rho(g)$ .

**Proposition 7.7.2.**

Let  $\|\cdot\|$  and  $\rho$  be respectively a norm and an homogeneous form adapted to a cone  $C$ . Let  $f, g \in C$  such that  $\rho(f) = \rho(g) \neq 0$  and satisfy

$$\|f - g\| \leq (e^{\theta_C(f,g)} - 1) \min(\|f\|, \|g\|). \quad (7.7.3)$$

*Proof.* Let  $f, g \in C$  such that  $\rho(f) = \rho(g) \neq 0$ . First, let's notice that the inequality is trivial if  $\theta_C(f, g) = \infty$ . Now, let assume that  $\theta_C(f, g) = \log \frac{\mu}{\lambda}$  where  $\mu < \infty$  and  $\lambda \neq 0$  and

$$\mu f - g \in C' \quad \text{and} \quad g - \lambda f \in C.$$

Then, by definition of the homogeneous form  $\rho$ , the following holds

$$\lambda\rho(f) \leq \rho(g) \leq \mu\rho(f),$$

and  $\lambda \leq 1 \leq \mu$ . Since  $\lambda$  is the smaller positive real number such that  $g - \lambda f \in C$ , it implies that

$$\begin{aligned}
&g - f - (\lambda - \mu)f \in C \\
&\text{and } (\mu - \lambda)f - (g - f) \in C.
\end{aligned}$$

Since the norm  $\|\cdot\|$  is adapted to a cone, then

$$\begin{aligned}
\|f - g\| &\leq (\mu - \lambda)\|f\| \\
&\leq \frac{\mu - \lambda}{\lambda}\|f\| \\
&\leq (e^{\theta_C(f,g)} - 1)\|f\|.
\end{aligned}$$

By exchanging the role of  $f$  and  $g$ , we obtain the result. □

We denote by  $\mathcal{C}^+(\Omega)$  the set of all continuous and positive functions on  $\Omega$  and we define  $\widetilde{\mathcal{C}^+(\Omega)} = \{f \in \mathcal{C}^+(\Omega) : \int f d\mu = 1\}$ .

**Proposition 7.7.3.**

Let  $f \in \mathcal{C}^+(\Omega)$  then  $\pi(g) = \{f \in \widetilde{\mathcal{C}^+(\Omega)}, \mathcal{O}_{\mathcal{C}^+(\Omega)}(f, g) < \infty\}$  is a complete metric space for the metric  $\mathcal{O}_{\mathcal{C}^+(\Omega)}$  associated to the cone of positive functions.

*Proof.* Let  $g \in \mathcal{C}^+(\Omega)$ ,  $\theta$  be the metric defined on  $\pi(g)$  and let  $(f_n)_{n \in \mathbb{N}}$  be a Cauchy sequence of  $\pi(g)$ . This sequence satisfies the following properties:

- By 7.7.2,

$$\|f_n - f_0\|_\infty \leq (e^{\mathcal{O}_C(f_n, f_0)} - 1) \|f_0\|_\infty$$

and  $\|f_n - f_{n+m}\|_\infty \leq (e^{\mathcal{O}_C(f_n, f_{n+m})} - 1) \|f_n\|_\infty.$

- $\mathcal{O}_C(f_n, f_0)$  is bounded and  $\mathcal{O}_C(f_n, f_{n+m})$  converges to 0 when  $n$  goes to infinity.

Thus,  $(f_n)_{n \in \mathbb{N}}$  is a Cauchy sequence for the uniform norm on  $\Omega$  and  $f$  its limit. Therefore,  $f \in \mathcal{C}^+(\Omega)$  because it's closed for the uniform norm and satisfies  $\int f d\mu = 1$  (since,  $\int f_n d\mu = 1$  for all  $n \in \mathbb{N}$ ). In the end, we show that  $f_n$  converges to  $f$  for the  $\theta$  metric. By definition of the projective metric, one has

$$\mu(f_n, f) = \sup_{x \in \Omega} \frac{f(x)}{f_n(x)}.$$

Since,  $f_n$  converges to  $f$  according to the uniform norm, then  $\mu(f_n, f)$  converges to 1. The same argument can be applied to prove that  $\lambda(f_n, f)$  converges to 1. Then,  $\mathcal{O}_C(f_n, f)$  goes to 0 and  $\pi(g)$  is a complete space.  $\square$

### 7.7.2 Proof V. Maume's theorem

We recall that the normalized operator  $\tilde{P}_\Phi$  is defined as follows (see 7.5.1).

$$\tilde{P}_\Phi f = \lambda_\Phi^{-1} h_\Phi^{-1} P_\Phi(f h_\Phi).$$

Thus

$$\tilde{P}_\Phi 1 = 1 \quad \text{and} \quad \tilde{P}_\Phi^* \mu_\Phi = \mu_\Phi.$$

We also recall that  $g$  is the inverse of the Jacobian of  $T$ , and  $g^{(k)}$  the inverse of the Jacobian of  $T^k$ , that is,

$$g = \frac{h_\Phi}{\lambda_\Phi h_\Phi \circ T} \exp(\phi) \quad \text{and} \quad g^{(k)} = \frac{h_\Phi}{\lambda_\Phi^k h_\Phi \circ T} \exp\left(\sum_{i=0}^{k-1} \phi \circ T^i\right).$$

(Of course  $g = g^{(1)}$ .) Therefore we have

$$\tilde{P}_\Phi f(x) = \sum_{T y = x} g(y) f(y) \quad \text{and} \quad \tilde{P}_\Phi^k f(x) = \sum_{T^k y = x} g^{(k)}(y) f(y).$$

Since  $h_\phi \in \mathcal{L}_\phi$ ,  $g^{(k)}$  satisfies the bounded distortion property: There exists  $K > 0$  such that for all  $k \in \mathbb{N}$ , and all  $x \stackrel{n}{\sim} y$  with  $n \geq 1$

$$\left| 1 - \frac{g^{(k)}(x')}{g^{(k)}(y')} \right| \leq K \text{var}_n \phi. \quad (7.7.4)$$

In fact,

$$\frac{g^{(k)}(x')}{g^{(k)}(y')} = \frac{h_\phi(x')}{h_\phi(y')} \frac{h_\phi(y)}{h_\phi(x)} e^{\sum_{i=0}^{k-1} \phi(T^i x') - \phi(T^i y)},$$

and

$$\begin{aligned} \frac{h_\phi(x')}{h_\phi(y')} &\leq \frac{\text{Lip}_\phi(h_\phi)}{\inf h_\phi} \text{var}_{n+k} \phi + 1, \\ \frac{h_\phi(y)}{h_\phi(x)} &\leq \frac{\text{Lip}_\phi(h_\phi)}{\inf h_\phi} \text{var}_n \phi + 1, \\ e^{\sum_{i=0}^{k-1} \phi(T^i x') - \phi(T^i y')} &\leq e^{\sum_{i=0}^{k-1} \text{var}_{n+i} \phi} \leq e^{k \text{var}_n \phi}. \end{aligned}$$

These inequalities imply the bounded distortion property (7.7.4). Since the measure  $\mu$  is mixing, for all finite partitions  $\mathcal{P}$  of  $\Omega$  defined by non-empty open sets, for all  $\alpha < 1 < \alpha'$ , there exists  $k_0$  such that  $\forall k > k_0$ ,

$$\forall A, B \in \mathcal{P}, \alpha \leq \frac{\mu(T^{-k} A \cap B)}{\mu(A)\mu(B)} \leq \alpha'. \quad (7.7.5)$$

**Remark 7.7.** For all  $A \in \mathcal{P}$ ,  $\mu(A) > 0$  because the support of  $\mu$  is  $\Omega$  and  $h$  is strictly positive.

Recall that a pattern  $p_s \in S^{\Lambda_s}$  determines a cylinder set  $[p_s] = \{\omega \in \Omega : \omega_{\Lambda_s} = p_s\}$ . For  $s \in \mathbb{N}^*$ , we denote by  $\mathcal{P}_s$  the finite partition of  $\Omega$  into cylinder set  $[p_s]$ . Let  $a, b$  be real numbers and  $\Lambda_{a,b}$  be the cone of functions which satisfies:

$$\begin{aligned} \forall A \in \mathcal{P}_s, 0 < \frac{1}{\mu(A)} \int_A f d\mu &\leq a \int f d\mu, \\ \text{Lip}_\phi(f) &\leq b \int f d\mu. \end{aligned}$$

We notice that  $\mathcal{C}_{a,b}$  is a good cone.

## Construction of cones

Now, let  $\ell \in \mathbb{N}$ . We construct a sequence of metrics  $d_\ell$ , a sequence of integers  $k_\ell$  and a sequence of cones  $\mathcal{C}_\ell$  such that  $\tilde{P}_\phi^{k_\ell} \mathcal{C}_\ell \subset \mathcal{C}_{\ell+1}$  and such that the diameter  $\Delta_\ell$  of  $\tilde{P}_\phi^{k_\ell} \mathcal{C}_\ell$  in  $\mathcal{C}_{\ell+1}$  is uniformly bounded in  $\ell$ .

Let's first study the action of  $\tilde{P}_\phi^k$  on the function of  $\mathcal{C}_{a,b}$ . We denote by  $D_0$  the diameter for the metric  $d_0$  of the partition  $\mathcal{P}_s$  defined as above,

$$D_0 = \sup_{A \in \mathcal{P}_s} \sup_{x, y \in P} d_0(x, y) = \text{var}_s \phi.$$

We will use the following notation, for  $a, b, c \in \mathbb{R}$ ,  $a = b \pm c$  means that  $b - c \leq a \leq b + c$ . We can prove easily that all functions  $\psi \in \mathcal{L}_\phi$  satisfy: For all  $A \in \mathcal{P}_s$  and all  $x \in P$ , one has

$$\frac{1}{\mu(A)} \int_A \psi d\mu - \text{Lip}_\phi(\psi)D_0 \leq \psi(x) \leq \frac{1}{\mu(A)} \int_A \psi d\mu + \text{Lip}_\phi(\psi)D_0. \quad (7.7.6)$$

Indeed, let  $x, y \in A$ , then

$$\psi(y) - \text{Lip}_\phi(\psi)d_0(x, y) \leq \psi(x) \leq \psi(y) + \text{Lip}_\phi(\psi)d_0(x, y).$$

Integrating over  $y$ , one has (7.7.6). Let  $f \in \mathcal{C}_{a,b}$ , and  $x \stackrel{n}{\sim} y$ ,  $n \geq 1$ ,

$$\begin{aligned} |\tilde{P}_\phi^k f(x) - \tilde{P}_\phi^k f(y)| &= \left| \sum_{T^k x' = x} f(x')g^{(k)}(x') - f(y')g^{(k)}(y') \right| \\ &\leq \sum_{T^k x' = x} g^{(k)}(x')|f(x') - f(y')| + \sum_{T^k x' = x} |g^{(k)}(x') - g^{(k)}(y')||f(y')| \\ &\leq \tilde{P}_\phi^k 1(x) \text{Lip}_\phi(f) \text{var}_{n+k}\phi + \sup |f| \sum_{T^k x' = x} g^{(k)}(x') \left| 1 - \frac{g^{(k)}(y')}{g^{(k)}(x')} \right| \\ &\leq \text{Lip}_\phi(f) \text{var}_{n+k}\phi + \sup |f| K \text{var}_n \phi. \end{aligned}$$

Thus, using Inequality (7.7.6) and the definition of  $\mathcal{C}_{a,b}$  we obtain

$$|\tilde{P}_\phi^k f(x) - \tilde{P}_\phi^k f(y)| \leq b \int f d\mu \left[ \text{var}_{n+k}\phi + \text{var}_n \phi K \frac{a + bD_0}{b} \right]. \quad (7.7.7)$$

For  $k_1 \in \mathbb{N}^*$  which will be fixed later on and  $D > 1$ , we consider

$$V_1(n) = D[\text{var}_{n+k_1}\phi + \text{var}_n].$$

The sequence  $(V_1(n))_{n \in \mathbb{N}}$  defines a new metric  $d_1$  on  $\Omega$  as follows:  $d_1(x, y) = V_1(n)$  if  $d(x, y) = \theta^n$ . Let  $\text{Lip}_\phi^{(1)}(f)$  be the Lipschitz constant for the metric  $d_1$ . Equation (7.7.7) shows that if  $D_0$  is sufficiently small and  $b$  sufficiently large then

$$\text{Lip}_\phi^{(1)}(\tilde{P}_\phi^{k_1} f) \leq \frac{b}{D} \int f d\mu.$$

On the other hand, for  $A \in \mathcal{P}_s$ , (7.7.6) implies:

$$\begin{aligned} \frac{1}{\mu(A)} \int_A \tilde{P}_\phi^k f d\mu &= \frac{1}{\mu(A)} \int_{T^{-k}A} f d\mu = \frac{1}{\mu(A)} \sum_B \int_{T^{-k}A \cap B} f d\mu \\ &= \sum_B \frac{\mu(T^{-k}A \cap B)}{\mu(A)\mu(B)} \int_B f d\mu \pm D_0 \text{Lip}_\phi(f) \sum_{B \in \mathcal{P}_1} \frac{\mu(T^{-k}A \cap B)}{\mu(A)\mu(B)} \mu(B). \end{aligned}$$

Thus, if  $k$  satisfies (7.7.5), we have:

$$[\alpha - \alpha' b D_0] \int f d\mu \leq \frac{1}{\mu(A)} \int_A \tilde{P}_\phi^k f d\mu \leq \alpha' [1 + b D_0] \int f d\mu. \quad (7.7.8)$$

Let  $\mathcal{C}_{a,b}^1$  the cone of functions of  $\mathcal{L}_\phi$  which satisfies:

$$\begin{aligned} \forall A \in \mathcal{P}_s, 0 < \frac{1}{\mu(A)} \int_A f d\mu &\leq a \int f d\mu, \\ \text{Lip}_\phi^{(1)}(f) &\leq b \int f d\mu. \end{aligned}$$

The following lemma shows that if the parameters are well-chosen,  $\tilde{P}_\phi^{k_1}$  is a contraction of  $\mathcal{C}_{a,b}$  into  $\mathcal{C}_1$ .

**Lemma 7.7.1.**

It exists  $D > 1$ ,  $k_1 \in \mathbb{N}^*$ ,  $a > 0$ ,  $b > 0$  and  $s \in \mathbb{N}^*$  such that,  $\tilde{P}_\phi^{k_1} \mathcal{C}_{a,b} \subset \mathcal{C}_{a,b}^1$ .  
 Moreover, the diameter  $\Delta_1$  of  $\tilde{P}_\phi^{k_1} \mathcal{C}_{a,b}$  in  $\mathcal{C}_{a,b}^1$  is bounded from above by  $2 \log \frac{D+1}{D-1}$ .

*Proof.* We fix  $0 < \zeta < 1$ ,  $\alpha < 1 < \alpha'$  and  $k_0$  such that (7.7.5) is satisfied for  $\alpha$  and  $\alpha'$ . Let

1.  $a$  be such that  $a \geq \zeta^{-1}(\alpha' + \alpha/2)$ ,
2.  $b$  be such that  $b > K(a + 1)$ ,
3.  $D > 1$  be such that  $\frac{D+1}{D-1} \geq \max \left[ \frac{2\alpha' + \alpha}{\alpha}, \frac{1}{1-\zeta} \right]$ ,
4.  $s \in \mathbb{N}$  be such that  $b(D + 1)\text{var}_s \phi < \frac{\alpha}{2\alpha'}$ ,
5.  $k_1$  be such that (7.7.5) is satisfied and  $D\text{var}_{s+k_1} \phi \leq \text{var}_s \phi$ .

Let  $f \in \mathcal{C}_{a,b}$  and assume that the previous conditions hold. Then,  $bD_0 = b\text{var}_s \phi < \frac{\alpha}{2\alpha'} < 1$ , Equation (7.7.8) implies:

$$\frac{\alpha}{2} \int \tilde{P}_\phi^k f d\mu \leq \frac{1}{\mu(A)} \int_A \tilde{P}_\phi^k f d\mu \leq \left( \alpha' + \frac{\alpha}{2} \right) \int \tilde{P}_\phi^k f d\mu \leq \zeta a \int \tilde{P}_\phi^k f d\mu \quad (7.7.9)$$

where  $a$  satisfies Condition 1 and Equation (7.7.7),  $\text{Lip}_\phi^{(1)}(\tilde{P}_\phi^k f) \leq \frac{b}{D} \int f d\mu$ .  
 Then,  $\tilde{P}_\phi^{k_1} \mathcal{C}_{a,b} \subset \mathcal{C}_{a,b}^1$ .

Now, we have to estimate the projective diameter. Let  $f, g \in \tilde{P}_\phi^{k_1} \mathcal{C}_{a,b}$  and  $\eta > 0$  such that  $\eta f - g \in \mathcal{C}_{a,b}^1$  and satisfies:

$$\forall A \in \mathcal{P}_s, 0 \leq \frac{\eta}{\mu(A)} \int_A f d\mu - \frac{1}{\mu(A)} \int_A g d\mu \leq a\eta \int_A f d\mu - a \int_A g d\mu,$$

and for  $x$  and  $y$  in the same cylinder set  $[p_1]$ ,

$$\begin{aligned} -b\eta \int f d\mu - b \int g d\mu &\leq \frac{\eta(f(x) - f(y)) - (g(x) - g(y))}{d_1(x, y)} \\ &\leq b\eta \int_A f d\mu - b \int_A g d\mu. \end{aligned}$$

To satisfy Condition 1, we need

$$\begin{aligned} \eta &\geq \sup_{A \in \mathcal{P}_s} \frac{a \int g d\mu - \frac{1}{\mu(A)} \int_A g d\mu}{a \int f d\mu - \frac{1}{\mu(A)} \int_A f d\mu} \\ \text{et } \eta &\geq \sup_{A \in \mathcal{P}_s} \frac{\int_A g d\mu}{\int_A f d\mu}. \end{aligned}$$

Since (7.7.9), we have for all  $A \in \mathcal{P}_s$ :

$$\begin{aligned} \frac{\int g d\mu - \frac{1}{\mu(A)} \int_A g d\mu}{\int f d\mu - \frac{1}{\mu(A)} \int_A f d\mu} &\leq \frac{1}{1-\zeta} \frac{\int g d\mu}{\int f d\mu} \\ \text{et } \frac{\int_A g d\mu}{\int_A f d\mu} &\leq \frac{2\alpha' + \alpha}{\alpha} \frac{\int g d\mu}{\int f d\mu}. \end{aligned}$$

Thus, to obtain Condition 1, it is enough that  $\eta$  satisfies:

$$\eta \geq \frac{\int g d\mu}{\int f d\mu} \max\left(\frac{1}{1-\gamma}, \frac{2\alpha' + \alpha}{\alpha}\right).$$

To satisfy Condition 2, we need, for  $x$  and  $y$  in the same cylinder set  $[p_1]$ ,

$$\begin{aligned} \eta &\geq \frac{b \int g d\mu - \frac{g(x)-g(y)}{d_1(x,y)}}{b \int f d\mu - \frac{f(x)-f(y)}{d_1(x,y)}} \\ \eta &\geq \frac{b \int g d\mu + \frac{g(x)-g(y)}{d_1(x,y)}}{b \int f d\mu + \frac{f(x)-f(y)}{d_1(x,y)}}, \end{aligned}$$

since  $\text{Lip}_\phi^{(1)}(f) \leq \frac{b}{D} \int f d\mu$  and  $\text{Lip}_\phi^{(1)}(g) \leq \frac{b}{D} \int g d\mu$ , it is enough to have:

$$\eta \geq \frac{\int g d\mu}{\int f d\mu} \frac{1+D}{D-1}.$$

In the same way, let  $\zeta > 0$  be such that  $g - \zeta f \in \mathcal{C}_{a,b}^1$ . It is enough that  $\zeta$  satisfies:

$$\zeta \leq \frac{\int g d\mu}{\int f d\mu} \min\left(1 - \gamma, \frac{D-1}{D+1}, \frac{\alpha}{2\alpha' + \alpha}\right).$$

Since  $D$  satisfies  $\frac{D+1}{D-1} \geq \frac{2\alpha'+\alpha}{\alpha}$  and  $\frac{D+1}{D-1} \geq \frac{1}{1-\zeta}$ , the diameter  $\Delta_1$  of  $P^k \mathcal{C}_{a,b}$  in  $\mathcal{C}_{a,b}^1$  is bounded by above by  $2 \log \frac{D+1}{D-1}$ .  $\square$

**Remark 7.8.** In the proof of Lemma 7.7.1, we used  $bD_0 < \frac{\alpha}{2\alpha'} < 1$ . Condition 5 implies that the diameter  $D_1 = V_1(s)$  of  $\mathcal{P}_s$  for the metric  $d_1$  is bounded by  $(D+1)\text{var}_s \phi$ . Condition 4 allows us to use a recurrence procedure. Indeed, we obtain  $bD_1 \leq \frac{\alpha}{2\alpha'}$ .

Fix  $a, b, D$  and  $s$  such that Lemma 7.7.1 is satisfied. Let

$$\begin{aligned} k_0 &= \inf\{k \in \mathbb{N} : \text{Lemma 7.7.1 is satisfied}\} \\ \text{et } k_1 &= \inf\{k > k_0 : D\text{var}_{k+s}\phi \leq \text{var}_s\phi\} \end{aligned}$$

We proceed in the same vein as (7.7.7). We obtain, for  $x$  and  $y$  such that  $x \sim^n y$ ,  $n \geq 1$  and  $f \in \mathcal{C}_{a,b}^1$ :

$$|\tilde{P}_\phi^k f(x) - \tilde{P}_\phi^k f(y)| \leq b \int f d\mu \left[ V_1(n+k) + \text{var}_n \phi K \frac{a+bD_1}{b} \right].$$

Thus, we define  $k_\ell$  and  $V_\ell(n)$  by the following induction:

- $k_{\ell+1} = \inf\{k > k_0 : DV_\ell(s+k_{\ell+1}) \leq \text{var}_s\phi\}$ ,
- the sequence  $(V_\ell(n))_{n \in \mathbb{N}}$  is defined by induction by

$$V_\ell(n) = D[V_{\ell-1}(n+k_\ell) + \text{var}_n \phi].$$

Each sequence  $(V_\ell(n))_{n \in \mathbb{N}}$  defines a metric on  $\Omega$  denoted by  $d_\ell$ . For  $f \in \mathcal{L}_\phi$ , we denote by  $\text{Lip}_\phi^{(\ell)}(f)$  the Lipschitz constant of  $f$  for the metric  $d_\ell$  and  $D_\ell$  the partition diameter of  $\mathcal{P}_s$  for the metric  $d_\ell$ . By construction,  $D_\ell = V_\ell(s) \leq$

$(D + 1)\text{var}_s\phi$  and thus,  $bD_\ell \leq \frac{\alpha}{2\alpha^\ell}$ . Let  $\mathcal{C}_{a,b}^\ell$  the cone of functions of  $\mathcal{L}_\phi$  which satisfies

$$\forall A \in \mathcal{P}_s, 0 < \frac{1}{\mu(A)} \int_A f d\mu \leq a \int f d\mu,$$

$$\text{Lip}_\phi^{(\ell)}(f) \leq b \int f d\mu.$$

Denote  $\mathcal{C}_{a,b}^0 = \mathcal{C}_{a,b}$ . Following the same path of the proof of Lemma 7.7.1, we obtain the following result.

**Proposition 7.7.4.**

The sequence of integers  $(k_\ell)_{\ell \in \mathbb{N}}$  and the sequence of cones  $(\mathcal{C}_{a,b}^\ell)$  satisfy:

- $\tilde{P}_\phi^{k_\ell} \mathcal{C}_{a,b}^{\ell-1} \subset \mathcal{C}_{a,b}^\ell, \ell \geq 1,$
- The diameter  $\Delta_\ell$  of  $\tilde{P}_\phi^{k_\ell} \mathcal{C}_{a,b}^{\ell-1}$  in  $\mathcal{C}_{a,b}^\ell$  is bounded by  $2 \log \frac{D+1}{D-1} := \Delta.$

Let  $\mathcal{C}$  be the cone of functions  $f \in \mathcal{L}_\phi$  such that  $0 < \sup_{x \in \Omega} |f(x)| \leq (a + 1) \int f d\mu$ . The following properties are derived directly from Definition 7.7.1 of  $\mathcal{C}$ .

1.  $\mathcal{C} \cap -\mathcal{C} = 0,$
2.  $\mathcal{C}$  is a convex cone,
3.  $\mathcal{C}$  is closed for the uniform norm topology.

In order to use this result, we need an adapted norm to  $\mathcal{C}$ . Let  $d > 0$ , we consider the norm

$$\|f\|_d = \max \left( d \left| \int f d\mu \right|, \|f\|_\infty \right).$$

**Remark 7.9.** For all  $d > 0$ , the norm  $\|\cdot\|_d$  is equivalent to the uniform norm on  $\Omega$ .

**Lemma 7.7.2.**

For all  $d \geq a + 1$ , the norm  $\|\cdot\|_d$  is adapted to the cone  $\mathcal{C}$ .

*Proof.* If  $f$  and  $g$  are such that  $f + g \in \mathcal{C}$  and  $f - g \in \mathcal{C}$ , then

$$f(x) - g(x) = \pm(a + 1) \int (f - g) d\mu$$

et

$$f(x) + g(x) = \pm(a + 1) \int (f + g) d\mu.$$

This implies the following

$$|g(x)| \leq (a + 1) \int f d\mu.$$

Moreover, if  $f - g \in \mathcal{C}$  and  $f + g \in \mathcal{C}$  then  $|\int g d\mu| \leq \int f d\mu$ . Thus, for  $x \in \Omega$  and  $d \geq a + 1$  one has,  $\|g\|_d \leq \|f\|_d$ .  $\square$

By the choice of  $a, b, s$  and  $k_\ell$ , the cones  $\mathcal{C}_{a,b}^\ell$  are sub-cones of  $\mathcal{C}$ . Indeed, if  $f \in \mathcal{C}_{a,b}^\ell$ ,  $f$  satisfies the following inequalities.

$$\begin{aligned} \forall a \in \mathcal{P}_s, 0 < \frac{1}{\mu(a)} \int_a f d\mu &\leq a \int f d\mu, \\ \text{Lip}_\phi^{(\ell)}(f) &\leq b \int f d\mu. \end{aligned}$$

Using the metric  $d_\ell$  instead of  $d_0$  in the proof of Proposition 7.7.4, we obtain

$$\begin{aligned} \sup |f| &\leq \sup \mathbb{E}(f|\mathcal{P}_s) + \text{Lip}_\ell(f) D_\ell, \\ \sup |f| &\leq a \int f d\mu + b D_\ell \int f d\mu, \\ \sup |f| &\leq (a+1) \int f d\mu \text{ car } b D_\ell \leq b(D+1) \text{var}_s \phi < 1. \end{aligned}$$

Proposition 7.7.1 on the cones implies that  $\mathcal{O}_\mathcal{C} \leq \mathcal{O}_{\mathcal{C}_{a,b}^\ell}$  for all  $\ell \in \mathbb{N}$ . Let  $\delta = \tanh \frac{\Delta}{4}$  and  $f \in \mathcal{C}_0$ ,  $P^{k_1+\dots+k_\ell} f \in \mathcal{C}_{a,b}^\ell$ , then one has,

$$\mathcal{O}_\mathcal{C}(\tilde{P}_\phi^{k_1+\dots+k_\ell} f, \mathbb{1}) \leq \mathcal{O}_{\mathcal{C}_{a,b}^\ell}(\tilde{P}_\phi^{k_1+\dots+k_\ell} f, \mathbb{1}) \leq \mathcal{O}_{\mathcal{C}_{a,b}^{\ell-1}}(\tilde{P}_\phi^{k_1+\dots+k_{\ell-1}} f, \mathbb{1}) \leq \dots \leq \delta^{\ell-1} \Delta.$$

For  $n \in \mathbb{N}$ , there exists a unique integer  $\ell(n)$  such that

$$k_1 + \dots + k_{\ell(n)} \leq n \leq k_1 + \dots + k_{\ell(n)+1}.$$

Thus, by Proposition 7.7.2 applied to  $\|\cdot\|_d$  and the homogeneous form  $\mu$ , we can estimate the speed of convergence of  $\tilde{P}^n f$  to  $\mu(f)$  for all  $f \in \mathcal{C}_0$ . Let  $n = k_1 + \dots + k_{\ell(n)} + u$  with  $u \geq 0$ , one has

$$\begin{aligned} \|\tilde{P}_\phi^n f - \mu(f)\|_d &= \|\tilde{P}_\phi^u [P^{k_1+\dots+k_{\ell(n)}} f - \mu(f)]\|_d \\ &\leq \|\tilde{P}_\phi^u\|_d \|\tilde{P}_\phi^{k_1+\dots+k_{\ell(n)}} f - \mu(f)\|_d \\ &\leq C_{te} \delta^{\ell(n)} \mu(f). \end{aligned} \tag{7.7.10}$$

The following lemma gives an estimation of the speed of convergence of  $\tilde{P}_\phi^n f$  to  $\mu(f)$  for all  $f \in \mathcal{L}_\phi$ .

**Lemma 7.7.3.**

*For all functions  $f \in \mathcal{L}_\phi$ , there exists  $R(f) \geq 0$  such that  $f + R(f)\mathbb{1} \in \mathcal{C}_0$  and  $R(f) \leq C_{te} \|f\|_{\mathcal{L}_\phi}$ .*

*Proof.* Let  $f \in \mathcal{L}_\phi$ .  $R(f)$  has to satisfy

$$\begin{aligned} R(f) &\geq \sup_{A \in \mathcal{P}_s} \frac{\frac{1}{\mu(A)} \int_A f d\mu - a \int f d\mu}{a-1}, \\ \text{and } R(f) &\geq \frac{\text{Lip}_\phi(f) - b \int f d\mu}{b}. \end{aligned}$$

Thus, we choose  $R(f)$  such that

$$R(f) = \max \left[ \frac{\text{Lip}_\phi(f)}{b} + \int f d\mu, \sup |f| \frac{a+1}{a-1} \right] \leq C_{te} \|f\|_{\mathcal{L}_\phi}.$$

□



Let  $f \in \mathcal{L}_\phi$  and  $f_{\mathcal{C}_0} = f + R(f)\mathbb{1} \in \mathcal{C}_0$ . The equation (7.7.10) and Lemma 7.7.2 imply

$$\begin{aligned} \|\tilde{P}_\phi^n f - \mu(f)\|_d &= \|\tilde{P}_\phi^n f_{\mathcal{C}_0} - \mu(f_{\mathcal{C}_0})\|_d + \|R(f)\tilde{P}_\phi^n \mathbb{1} - R(f)\mu(1)\|_d \\ &\leq C_{te}\delta^{\ell(n)}\|f\|_{\mathcal{L}_\phi}. \end{aligned}$$

Since the norm  $\|\cdot\|_d$  is equivalent to the uniform norm on  $\Omega$ , we obtain the V. Maume's result. For all functions  $f \in \mathcal{L}_\phi$ ,

$$\|\tilde{P}_\phi^n f - \mu(f)\|_\infty \leq C_{te}\delta^{\ell(n)}\|f\|_{\mathcal{L}_\phi}.$$

# Chapter 8

## Probabilistic cellular automata

### 8.1 Introduction

Probabilistic cellular automata (PCA) are Markov chains on shift spaces, such that the probability transition rule is local and shift-invariant. These models have been extensively studied as they represent a paradigm of complex spatiotemporal behavior. They can be seen as both, stochastic perturbations of deterministic cellular automata (CA), as well as lattice statistical mechanics models. This last point of view, allowing the use of techniques from statistical mechanics, has led to results concerning the so called high-noise regime, where ergodicity and decay of correlations can be derived from results in high-temperature regimes for the corresponding statistical mechanical model. For further details on PCA seen as Gibbs measures on histories (in space-time) we mention [40]. This approach has also furnished examples of loss of ergodicity via phase transitions in the corresponding statistical mechanical model. The high-noise regime has been mainly studied using the Dobrushin approach for high temperature statistical mechanical models, leading to progressively finer and more explicit conditions on the noise strength, ensuring ergodicity, decay of correlations and the large deviation property. A few alternative approaches have been tried, among which the coupling method. In this chapter we prove that, under the classical Dobrushin-Shlosman conditions, the evolution is a contraction in  $\bar{d}$ -distance which ensures uniform exponential ergodicity. In a second part we prove that under those conditions, the spatio-temporal process defined by the PCA has the Gaussian Concentration Bound for space-time observables whose Lip-vector has bounded  $\ell_2$  norm.

### 8.2 Setting

The configurations of our PCA belong to the product space  $\Omega := S^{\mathbb{L}}$ . We recall that the shift action  $(T_x, x \in \mathbb{L})$  is defined as: for each  $x \in \mathbb{L}$ ,  $T_x : \Omega \rightarrow \Omega$  and  $(T_x \omega)_y = \omega_{y-x}$ , for all  $y \in \mathbb{L}$ . Let us denote by  $\mathfrak{B}$  the Borel sigma-algebra on  $\Omega$ . For  $a \in S^\Lambda$ , with  $[a]$  we denote the cylinder set  $\{\eta \in \Omega : \eta_\Lambda = a\}$ .

A PCA is a discrete-time Markov process  $(\omega^t)_{t \in \mathbb{N}}$  on  $\Omega$ , defined by a probability kernel  $\Phi : \mathfrak{B} \times \Omega \rightarrow [0, 1]$  (the definition of probability kernel can be found in [39] for instance) with this structure: there exists  $0 \in U \Subset \mathbb{L}$  (a finite subset containing 0), and a probability transition function  $\phi : S \times S^U \rightarrow [0, 1]$

such that for each  $\Lambda \subset \mathbb{L}$  finite,  $a \in S^\Lambda$ , and  $\omega \in \Omega$  we have

$$\Phi([a], \omega) = \prod_{x \in \Lambda} \phi(a_x, \omega_{U+x}). \quad (8.2.1)$$

Notice that  $\phi(\cdot, \eta_U)$  is a probability vector in  $[0, 1]^S$ .

A deterministic cellular automata (CA) is a  $T$ -invariant continuous transformation  $\bar{\Phi} : \Omega \rightarrow \Omega$ , which by virtue of the Curtis-Hedlund-Lyndon theorem is necessarily defined by a local transition rule  $\bar{\phi} : S^U \rightarrow S$  in the following way. For each  $x \in \mathbb{L}$  we have

$$\left(\bar{\Phi} \omega\right)_x = \bar{\phi}(\omega_U). \quad (8.2.2)$$

A CA can be seen as a degenerated PCA by considering a probability transition function of the kind  $\phi(a_0, b_U) = \delta_{\bar{\phi}(b_U)}^{a_0}$ , with  $\delta_b^a$  denoting the Kronecker delta.

The space-time configurations of the PCA belong to the product space  $X := \Omega^{\mathbb{N}}$  for which we consider the product topology and the Borel sigma-algebra  $\mathfrak{F}$  generated by the sets in  $\mathfrak{B} \times \mathfrak{B}$ . Let  $\mathcal{M}^1(X)$  and  $\mathcal{M}^1(\Omega)$  be the Banach spaces of signed measures in  $X$  and  $\Omega$  respectively, and let  $\mathcal{M}(X)$  and  $\mathcal{M}(\Omega)$  be the corresponding simplices of probability measures. Given  $F \in \mathbb{L} \times \mathbb{N}$ , and  $a \in S^F$ , with  $[a]$  we will denote the corresponding space-time cylinder  $\{\omega \in X : \omega_s^t = a_s^t, \forall (s, t) \in F\}$ . With  $a_F$  we denote the restriction of  $a$  to the coordinates of  $F$ . When  $F = \Lambda \times \{t\} \subset \mathbb{L} \times \mathbb{N}$ ,  $a_F$  will also be denoted by  $a_\Lambda^t$ .

A particular class of cylinders we will distinguish are the one defined by space-time funnels. For  $\Lambda \in \mathbb{L}$  and  $\mathcal{T} \in \mathbb{N}$ , let

$$F_\Lambda^{\mathcal{T}} = \cup_{t=0}^{\mathcal{T}} (\Lambda + U_{\mathcal{T}-t}) \times \{t\}, \quad (8.2.3)$$

where  $U_\tau$  is defined recursively as follows:  $U_0 = U$ , and for each  $\tau \geq 0$ ,  $U_{\tau+1} = U_\tau + U$ . The set  $F_\Lambda^{\mathcal{T}}$  is the time- $\mathcal{T}$  funnel with base  $\Lambda$ . The funnel  $F_\Lambda^{\mathcal{T}}$  contains all the sites in space-time, required to determine the distribution of configurations  $\Lambda \in \mathbb{L}$  at time  $\mathcal{T}$ . When the base of the funnel is a singleton  $\Lambda = \{x\}$ , we will use the notation  $F_x^{\mathcal{T}}$  to simplify. Cylinder sets supported by space-time funnels suffice to generate all the measurable sets in  $\mathfrak{F}$ .

The opposite structure of a funnel is a light cone. The light cone emerging from  $(x, t) \in \mathbb{L} \times \mathbb{N}$ , is the set

$$\mathcal{L}_{(x,t)} := \cup_{\tau=0}^{\infty} (x + U_\tau) \times \{t + \tau\}. \quad (8.2.4)$$

Given initial probability distribution  $\mu \in \mathcal{M}(\Omega)$ , the PCA defines a space-time distribution  $\mathbb{P}_\mu \in \mathcal{M}(X)$  as follows. Let  $\Lambda \in \mathbb{L}$ ,  $\mathcal{T} \in \mathbb{N}$  and  $a \in S^{F_\Lambda^{\mathcal{T}}}$  then

$$\begin{aligned} \mathbb{P}_\mu[a] &= \mu \left[ a_{\Lambda+U_{\mathcal{T}}}^0 \right] \prod_{t=1}^{\mathcal{T}} \Phi \left( \left[ a_{\Lambda+U_{\mathcal{T}-t}}^t \right], \omega^{t-1} \right) \\ &= \mu \left[ a_{\Lambda+U_{\mathcal{T}}}^0 \right] \prod_{t=1}^{\mathcal{T}} \prod_{x \in \Lambda + U_{\mathcal{T}-t}} \phi \left( a_x^t, a_{x+U}^{t-1} \right), \end{aligned} \quad (8.2.5)$$

for any choice  $\omega^t \in \left[ a_{\Lambda+U_{\mathcal{T}-t}}^t \right]$ .

The PCA induces the transformation  $\mathcal{P}_\Phi : \mathcal{M}(\Omega) \rightarrow \mathcal{M}(\Omega)$  given by

$$\mathcal{P}_\Phi \nu (B) = \int_X \Phi(B, \omega) d\nu(\omega), \quad (8.2.6)$$

for all  $B \in \mathfrak{B}$ . Since  $\Phi$  is  $T$ -invariant (recall that it means that  $\Phi(T_x^{-1}A, \omega) = \Phi(A, T_x\omega)$  for each  $A \in \mathfrak{B}$ ,  $\omega \in \Omega$  and  $x \in \mathbb{L}$ ), then  $\mathcal{P}_\Phi$  preserves the simplex  $\mathcal{M}_T(\Omega)$  of  $T$ -invariant probability measures. Since  $\mathcal{P}_\Phi$  is continuous in the weak- $*$  topology, the Schauder fixed-point theorem ensures the existence of at least one stationary measure, *i.e.*, a measure  $\mu \in \mathcal{M}_T(\Omega)$  such that  $\mathcal{P}_\Phi \mu = \mu$ . The PCA is said to be ergodic if there exists a unique attractive probability measure  $\mu^* \in \mathcal{M}(\Omega)$ , *i.e.*, such that for each probability measure  $\mu \in \mathcal{M}(\Omega)$ ,  $\mathcal{P}_\Phi^\tau \mu \rightarrow \mu^*$  in the weak- $*$  topology. Evidently  $\mu^*$  has to be  $T$ -invariant. If  $\mathcal{P}_\Phi$  happens to be a contractive map with respect to a distance  $D$  compatible with the weak- $*$  topology, Banach's contracting map theorem ensures ergodicity. Furthermore, in this case the PCA is not only ergodic, but uniformly ergodic, which means that  $\mathcal{P}_\Phi^\tau \mu[a] \rightarrow \mu_\infty[a]$  for each  $a \in S$ , uniformly on  $\mu$ . It is unknown whether uniform ergodicity is a stronger condition than simple ergodicity.

The PCA also defines a non-negative linear action  $P_\Phi$  on the set of continuous functions  $\mathcal{C}(X)$ ,

$$(P_\Phi K)(\omega) := \int_X K(\eta) \Phi(d\eta, \omega). \quad (8.2.7)$$

Exponential convergence of  $\mathcal{P}_\Phi \mu$  in distance  $D$  does not ensure that the unique stationary distribution  $\mu^*$  is ergodic with respect to  $T$ , much less mixing [59]. Mixing requires more than just  $D$ -contractiveness. For instance, if for some  $f : \mathbb{N} \rightarrow [0, \infty)$  such that  $\lim_{\tau \rightarrow \infty} f(\tau) = 0$ , and for some product measure  $\mu \in \mathcal{M}(\Omega)$  we have

$$|\mathcal{P}_\Phi^\tau \mu[a] - \mu^*[a]| \leq |\Lambda| f(\tau), \quad \forall \Lambda \in \mathbb{L}, a \in S^\Lambda \text{ and } \tau \in \mathbb{N}, \quad (8.2.8)$$

then the unique stationary measure  $\mu^*$  has to be mixing with respect to  $T$ . This is the way mixing of  $\mu^*$  is proved in [57, Corollary 1]. Proposition 2.1 in [59] synthesizes the rationale behind this approach. This kind of uniform convergence is obtained for instance when  $\mathcal{P}_\Phi$  is a contraction with respect to the  $\bar{d}$ -distance, a distance finer than  $D$ .

As mentioned in the previous paragraph, the ideal situation arrives when  $\mathcal{P}_\Phi$  is a contraction with respect to the  $\bar{d}$ -distance. We recall that this distance is defined as follows. Given two probability measures  $\mu, \nu \in \mathcal{M}_T^1(\Omega)$ , a coupling between them is a probability measure  $\lambda \in \mathcal{M}(\Omega \times \Omega)$  such that  $\lambda(B \times \Omega) = \mu(B)$  and  $\lambda(\Omega \times B) = \nu(B)$  for each  $B \in \mathfrak{B}$ . Let us denote by  $\mathcal{J}_T(\mu, \nu)$  be the set of all the  $T$ -invariant couplings between the  $\mu, \nu \in \mathcal{M}_T^1(\Omega)$ , then  $\bar{d} : \mathcal{M}_T^1(\Omega) \times \mathcal{M}_T^1(\Omega) \rightarrow [0, 1]$  given by

$$\bar{d}(\mu, \nu) = \inf_{\lambda \in \mathcal{J}_T(\mu, \nu)} \lambda \left( \bigcup_{a_0 \neq b_0} [a_0] \times [b_0] \right). \quad (8.2.9)$$

This distance makes  $\mathcal{M}_T^1(\Omega)$  a complete, non-compact, metric space. In this topology, limits of converging sequences of ergodic measures are ergodic. Similarly for sequences of mixing measures. Furthermore, entropy is continuous with respect to this topology (see [78]). It is not hard to prove that whenever  $\mathcal{P}_\Phi$  is a contraction in distance  $\bar{d}$ , then the PCA defined by  $\mathcal{P}_\Phi$  is uniformly

exponentially convergent, i.e., Inequality (8.2.8) holds with  $f(\tau) = \exp(c - \gamma \tau)$  for some  $c$  and  $\gamma > 0$ .

It is worth mentioning that, according to [56, Theorem 1], for the class of attractive PCA i.e., for all increasing functions  $K$ ,  $P_\Phi K$  is still increasing (which is a notion related to the attractiveness of potentials in the theory of Gibbs/g-measures),  $T$ -mixing of the unique PCA stationary distribution  $\mu^*$  implies (8.2.8) with  $f(\tau) \rightarrow 0$  exponentially fast.

The uniform exponential convergence (Equation (8.2.8) with  $f(\tau) = \exp(c - \gamma \tau)$  for some  $c$  and  $\gamma > 0$ ), when it happens, ensures a wealth of nice limit properties for the process  $\mathbb{P}_\mu$ , and ensures the Gaussian concentration bound. And conversely, non-ergodicity or even not fast enough convergence towards the unique stationary measure leads to the violation of the Gaussian concentration bound.

Let us recall these notions in this context. Let  $K : X \rightarrow \mathbb{R}$  be a continuous function and  $(x, t) \in \mathbb{L} \times \mathbb{N}$ . We denote by

$$\Delta_{(x,t)}(K) := \sup \left\{ |K(\omega) - K(\eta)| : \omega_{(s,\tau)} = \eta_{(s,\tau)}, \forall (s,\tau) \neq (x,t) \right\} \quad (8.2.10)$$

the spatio-temporal oscillation of  $K$  at  $(x, t)$ . An observable  $K : X \rightarrow \mathbb{R}$  is local if there exists  $\Lambda_K \in \mathbb{L} \times \mathbb{N}$  such that for all  $\omega, \omega', \omega'' \in X$ ,  $K(\omega_{\Lambda} \omega'_{\Lambda^c}) = K(\omega_{\Lambda} \omega''_{\Lambda^c})$ . Let's denote by  $\Lambda_K := \text{supp}(K)$  the support of  $K$ . We will refer to observables  $K : X \rightarrow \mathbb{R}$  satisfying

$$\|\Delta(K)\|_2^2 := \sum_{(x,t) \in \mathbb{L} \times \mathbb{N}} \Delta_{(x,t)}^2 K < \infty, \quad (8.2.11)$$

as Lip-bounded and will denote by  $\mathcal{L}_B$  the set of all Lip-bounded observables. Clearly the set of local observables forms a dense set, with respect to the sup-norm, inside  $\mathcal{L}_B$ , which in its turn is dense inside the Banach space  $\mathcal{C}(X)$  of continuous observables. We recall what we mean by Gaussian concentration bound.

**Definition 8.2.1.**

Let  $\mathbb{P}_\mu \in \mathcal{M}(X)$  be a space-time distribution with an initial probability measure  $\mu \in \mathcal{M}(\Omega)$ . We say that the PCA satisfies a Gaussian concentration bound if there exists  $D = D(\mathbb{P}_\mu) > 0$  such that for all functions  $K \in \mathcal{L}_B$ , we have

$$\mathbb{E}[\exp(K - \mathbb{E}[K])] \leq \exp\left(D \|\Delta(K)\|_2^2\right). \quad (8.2.12)$$

The expectation in (8.2.12) is taken with respect to the process  $\mathbb{P}_\mu$ , and should hold for each initial distribution  $\mu \in \mathcal{M}(\Omega)$ .

Summarizing, if  $\mathcal{P}_\Phi : \mathcal{M}(\Omega) \rightarrow \mathcal{M}(\Omega)$  has a unique stationary probability measure  $\mu^*$ , which necessarily is  $T$ -invariant, and if  $\mu^*$  uniformly attracts (in the sense of contraction) all initial distributions as in (8.2.8), then  $\mu^*$  is mixing with respect to  $T$ . If this convergence is exponential, then the space-time measure  $\mathbb{P}_\mu$  defined by the PCA and an initial distribution  $\mu \in \mathcal{M}(\Omega)$  satisfies what we call the space-time Gaussian Concentration Bound (8.2.12). A standard way to obtain this convergence is to ensure that the transition probability  $\phi$  satisfies a condition ensuring the contractiveness of the operator  $P_\phi$ . As mentioned above, under this kind of high-noise conditions, limit properties such as the Central Limit Theorem and the Large Deviation Principle hold.

These limit properties have been already established in previous works. In the present chapter we will be mainly concerned with the GCB.

### 8.3 High-noise regime

The high-noise regime corresponds to a situation where the transition probability function  $\phi : S \times S^U \rightarrow [0, 1]$  is so close to equidistribution that the PCA is a contraction in the  $\bar{d}$ -distance and therefore satisfies uniform exponential convergence. The first and most natural condition ensuring this contractiveness is the  $C_V$  condition by Dobrushin and Shlosman (see [28]). Taking into account that in the literature the computations usually refer to the action of the PCA on regular functions (either local or having finite Lip norm), we present our own formulation of the contractiveness result, referred to the  $\bar{d}$ -distance. It is nevertheless equivalent and the proof follows the standard reasoning: The convergence in  $\bar{d}$ -topology could be deduced from the fact that, under Dobrushin's type conditions as those used in [52, 57, 79], the action of  $\mathcal{P}$  is a contraction in  $\mathcal{M}_T(S^{\mathbb{Z}})$  with respect to the Wasserstein distance as it is presented in [48].

#### Theorem 8.3.1.

Let  $\mathcal{P}_\Phi$  be a PCA with transition probability function  $\phi : S \times S^U \rightarrow [0, 1]$ .  
If

$$\gamma_\phi := \sum_{x \in U} \max \left\{ \|\phi(\cdot, c) - \phi(\cdot, d)\|_{\text{TV}} : c, d \in S^U, c_s = d_s \forall s \neq x \right\} < 1,$$

then  $\bar{d}(\mathcal{P}_\phi \mu, \mathcal{P}_\phi \nu) \leq \gamma_\phi \bar{d}(\mu, \nu)$  for all  $\mu, \nu \in \mathcal{M}_T^1(\Omega)$ .

Recall that  $\|p - q\|_{\text{TV}} := 1/2 \sum_{a \in S} |p(a) - q(a)| = \sum_{a \in A} (p - q)^+(a)$  is the total variation distance between the probability vectors  $p$  and  $q \in S^{[0,1]}$ . This theorem holds for any lattice  $\mathbb{L}$  we consider, and ensures the following.

#### Corollary 8.3.1.

Under the hypotheses of Theorem 8.3.1, the PCA defined by  $\Phi$  is uniformly exponentially ergodic, i.e., there exists a unique  $\mu^* \in \mathcal{M}_T^1(\Omega)$  such that  $\mathcal{P}_\Phi \mu^* = \mu^*$  and

$$|\mathcal{P}_\Phi^t \mu[a] - \mu^*[a]| \leq \frac{|\Lambda|}{1 - \gamma_\phi} \gamma_\phi^t,$$

for each  $\Lambda \in \mathbb{L}$  and  $a \in S^\Lambda$ . Furthermore  $\mu^*$  is mixing.

*Proof.* Under the hypotheses of Theorem 8.3.1,  $\mathcal{P}_\phi$  is a contraction in  $\bar{d}$ -distance. The set  $\mathcal{M}_T^1(\Omega)$  with the  $\bar{d}$ -topology is complete. Banach's fixed point theorem ensures the existence and uniqueness of  $\mu^* \in \mathcal{M}_T^1(\Omega)$  such that  $\mathcal{P}_\Phi \mu^* = \mu^*$ . Furthermore, the theorem establishes that

$$\bar{d}(\mathcal{P}_\phi^t \mu, \mu^*) \leq \frac{\bar{d}(\mathcal{P}_\phi \mu, \mu)}{1 - \gamma_\phi} \gamma_\phi^t.$$

Clearly  $\bar{d}(\mathcal{P}_\Phi \mu, \mu) \leq 1$ , since  $\bar{d}(\mathcal{P}_\Phi \mu, \mu) \leq \sum_{a,b \in S, a \neq b} \lambda([a] \times [b])$  for each coupling  $\lambda \in \mathcal{J}_T(\mathcal{P}_\Phi \mu, \mu)$ . The coupling inequality (see [55] for details) establishes that for each  $\Lambda \in \mathbb{L}$

$$\max_{a \in S^\Lambda} |\mathcal{P}_\Phi^t \mu[a] - \mu^*[a]| \leq \sum_{a,b \in S^\Lambda, a \neq b} \lambda([a] \times [b]),$$

for each coupling  $\lambda \in \mathcal{J}_T(\mathcal{P}_\Phi^t \mu, \mu^*)$ . Taking into account that

$$\sum_{a,b \in S^\Lambda, a \neq b} \lambda([a] \times [b]) \leq \sum_{x \in \Lambda} \sum_{a_x \neq b_x} \lambda([a_x] \times [b_x]),$$

optimizing over the  $T$ -invariant couplings, we obtain

$$\max_{a \in S^\Lambda} |\mathcal{P}_\Phi^t \mu[a] - \mu^*[a]| \leq \sum_{x \in \Lambda} \bar{d}(\mathcal{P}_\Phi^t \mu, \mu^*) \leq \frac{|\Lambda|}{1 - \gamma_\phi} \gamma_\phi^t.$$

Finally, since in  $\bar{d}$ -distance converging limits of mixing measures are mixing (see [78] for details), taking  $\mu \in \mathcal{M}_T^1(\Omega)$  a mixing measure, necessarily  $\mu^* = \lim_{t \rightarrow \infty} \mathcal{P}_\Phi^t \mu$  is mixing. Alternatively we can use the previous inequality, which establishes uniform exponential ergodicity of the PCA, to directly prove the mixing. Indeed, by taking  $\mu \in \mathcal{M}_T^1(\Omega)$  a product measure and  $t \in \mathbb{N}$  and  $\|x\|_\infty > t \text{diam}(U) + \text{diam}(\Lambda)$ , we have

$$\begin{aligned} |\mu^*([a] \cap T_x[b]) - \mathcal{P}_\Phi^t \mu([a] \cap T_x[b])| &\leq \frac{2|\Lambda| \gamma_\phi^t}{1 - \gamma_\phi} \\ \mathcal{P}_\Phi^t \mu([a] \cap T_x[b]) &= (\mathcal{P}_\Phi^t \mu[a])(\mathcal{P}_\Phi^t \mu[b]) \\ |(\mathcal{P}_\Phi^t \mu[a])(\mathcal{P}_\Phi^t \mu[b]) - \mu^*[a] \mu^*[b]| &\leq \frac{|\Lambda| \gamma_\phi^t}{1 - \gamma_\phi} (\mu^*[a] + \mu^*[b]) + \frac{|\Lambda|^2 \gamma_\phi^{2t}}{1 - \gamma_\phi}. \end{aligned}$$

Hence, for all  $\|x\|_\infty$  sufficiently large, we have

$$|\mu^*([a] \cap T_x[b]) - \mu^*[a] \mu^*[b]| \leq \frac{4|\Lambda| \gamma_\phi^{\|x\|_\infty / \text{diam}(U)}}{(1 - \gamma_\phi) \gamma_\phi^{(\text{diam} \Lambda + 1) / \text{diam}(U)}}.$$

□

As we show in the next section, under the conditions of Theorem 8.3.1, the process has space-time GCB.

**Proof of Theorem 8.3.1.** To ease the notation, for each  $c \in S^U$  let us denote by  $\phi_c$  the probability vector  $\phi(\cdot, c)$ . Now, let  $\mathcal{P}_{\bar{\phi}}$  be the PCA on  $\Omega \times \Omega$  defined by the local transition probability function  $\bar{\phi} : (S \times S^U) \times (S \times S^U) \rightarrow [0, 1]$  such that

$$\bar{\phi}(a, c; b, d) = \begin{cases} \frac{(\phi_c - \phi_d)^+(a) (\phi_d - \phi_c)^+(b)}{\|\phi_c - \phi_d\|_{\text{TV}}} + \delta_a^b \phi_c \wedge \phi_d(a) & \text{if } \phi_c \neq \phi_d, \\ \delta_a^b \phi(a, c) & \text{if } \phi_c = \phi_d, \end{cases} \quad (8.3.1)$$

where, as usual,  $p^+$  denotes the positive part of  $p$  while  $p \wedge q = \min(p, q)$ . This transition probability function is nothing but the maximal coupling between  $\phi_c$  and  $\phi_d$  seen as probability distributions. It is easy to verify that

$$\sum_{a \in S} \bar{\phi}(a, c; b, d) = \phi(b, d), \text{ and } \sum_{b \in S} \bar{\phi}(a, c; b, d) = \phi(a, c),$$

which ensures that  $\bar{\phi}$  correctly defines a probability kernel. Now, let  $\lambda \in \mathcal{J}_T(\mu, \nu)$  and consider its image  $\lambda' = \mathcal{P}_{\bar{\phi}}\lambda$ . For each  $\Lambda \in \mathbb{L}$  and  $a \in S^\Lambda$  we have,

$$\begin{aligned} \lambda'([a] \times X) &= \sum_{b \in S^\Lambda} \sum_{c, d \in S^{\Lambda+U}} \lambda([c] \times [d]) \prod_{x \in \Lambda} \bar{\phi}(a_x, c_{x+U}; b_x, d_{x+U}) \\ &= \sum_{c, d \in S^{\Lambda+U}} \lambda([c] \times [d]) \prod_{x \in \Lambda} \left( \sum_{b_x \in S} \bar{\phi}(a_x, c_{x+U}; b_x, d_{x+U}) \right) \\ &= \sum_{c, d \in S^{\Lambda+U}} \lambda([c] \times [d]) \prod_{x \in \Lambda} \phi(a_x, c_{x+U}) = \sum_{c \in S^{\Lambda+U}} \mu([c]) \prod_{x \in \Lambda} \phi(a_x, c_{x+U}) \\ &= \mathcal{P}_{\Phi}\mu[a]. \end{aligned}$$

Since  $\bar{\phi}$  is symmetric, following the analogous computation we obtain  $\mathcal{P}_{\bar{\phi}}(X \times [b]) = \mathcal{P}_{\Phi}\nu[b]$  for each  $b \in S^{\Lambda+U}$ . Hence,  $\lambda' \in \mathcal{J}_T(\mathcal{P}_{\Phi}\mu, \mathcal{P}_{\Phi}\nu)$ , and therefore

$$\begin{aligned} \bar{d}(\mathcal{P}_{\Phi}\mu, \mathcal{P}_{\Phi}\nu) &\leq \sum_{a, b \in S, a \neq b} \lambda'([a] \times [b]) = \sum_{c, d \in S^U} \sum_{a \neq b} \bar{\phi}(a, c; b, d) \lambda([c] \times [d]) \\ &= \sum_{c \neq d} \frac{\lambda([c] \times [d])}{\|\phi_c - \phi_d\|_{\text{TV}}} \sum_{a \in S} (\phi_c - \phi_d)^+(a) \sum_{b \neq a} (\phi_d - \phi_c)^+(b) \\ &= \sum_{c \neq d} \frac{\lambda([c] \times [d])}{\|\phi_c - \phi_d\|_{\text{TV}}} \sum_{a \in S} ((\phi_c - \phi_d)^+(a)) \left( \|\phi_c - \phi_d\|_{\text{TV}} - (\phi_d - \phi_c)^+(a) \right) \\ &\leq \sum_{c \neq d} \lambda([c] \times [d]) \|\phi_c - \phi_d\|_{\text{TV}}. \end{aligned}$$

Let  $n : \{1, 2, \dots, |U|\} \rightarrow U$  be a numbering of the coordinates in  $U$ . For each  $c, d \in S^U$  with  $c \neq d$  we have

$$\begin{aligned} \lambda([c] \times [d]) \|\phi_c - \phi_d\|_{\text{TV}} &= \lambda([c] \times [d]) \sum_{k=1}^{|U|} \|\phi(\cdot, c_{n \leq n(k)} d_{n > n(k)}) - \phi(\cdot, c_{n < n(k)} d_{n \geq n(k)})\|_{\text{TV}}, \end{aligned}$$

where  $c_{n \leq n(k)} d_{n > n(k)} \in S^U$  is the configuration whose first  $k$  coordinates coincide with  $c$  and the rest with  $d$ . Similarly for  $c_{n < n(k)} d_{n \geq n(k)}$ . Notice that  $c_{n \leq n(k)} d_{n > n(k)}$  and  $c_{n < n(k)} d_{n \geq n(k)}$  differ at most in the site  $n(k)$ . With this, and taking into account that  $\lambda$  is  $T$ -invariant, we have

$$\begin{aligned} \sum_{c \neq d} \lambda([c] \times [d]) \|\phi_c - \phi_d\|_{\text{TV}} &\leq \sum_{c \neq d} \lambda([c] \times [d]) \sum_{c_{n(k)} \neq d_{n(k)}} \max\{\|\phi_c - \phi_d\|_{\text{TV}} : c_s = d_s, s \neq n(k)\} \\ &\leq \sum_{k=1}^{|U|} \max\{\|\phi_c - \phi_d\|_{\text{TV}} : c_s = d_s, s \neq n(k)\} \sum_{c, d: c_{n(k)} \neq d_{n(k)}} \lambda([c] \times [d]) \\ &= \gamma_\phi \sum_{c, d: c_{n(k)} \neq d_{n(k)}} \lambda([c] \times [d]) \\ &= \gamma_\phi \sum_{a, b \in S, a \neq b} \lambda([a] \times [b]). \end{aligned}$$

Since the inequality holds for each  $\lambda \in \mathcal{J}_T(\mu, \nu)$ , we finally obtain

$$\bar{d}(\mathcal{P}_{\Phi}\mu, \mathcal{P}_{\Phi}\nu) \leq \gamma_\phi \inf \left\{ \sum_{a, b \in A, a \neq b} \lambda([a] \times [b]) : \lambda \in \mathcal{J}_T(\mu, \nu) \right\} = \gamma_\phi \bar{d}(\mu, \nu).$$



□

## 8.4 Gaussian Concentration Bound

In the regime of high noise the PCA has a Gaussian concentration bound. We have the following.

### Theorem 8.4.1.

Let  $\mathcal{P}_\Phi$  be a PCA whose transition probability function  $\phi : S \times S^U \rightarrow [0, 1]$  is such that  $\phi(a, b) > 0$  for all  $a \in S$  and  $b \in S^U$ , and such that

$$\gamma_\phi := \sum_{x \in U} \max \left\{ \|\phi(\cdot, c) - \phi(\cdot, d)\|_{\text{TV}} : c, d \in S^U, c_s = d_s \forall s \neq x \right\} < 1.$$

Then, for each  $\mu \in \mathcal{M}(\Omega)$  the PCA satisfies a Gaussian concentration bound.

In the proof we will use the classical Azuma-Hoeffding approach, which we think is the more appropriate for this case. We will also make use of the following lemma.

### Lemma 8.4.1.

Let  $\mu \in \mathcal{M}(\Omega)$  and for each  $a \in S$  and  $x \in \mathbb{L}$ , let  $\mu_a^x := \mu(\cdot | \omega_x = a) \in \mathcal{M}(\Omega)$ . Under the hypotheses of Theorem 8.3.1 there exists a coupling  $\lambda_{a,b}^x \in \mathcal{J}_T(\mathbb{P}_{\mu_a^x}, \mathbb{P}_{\mu_b^x})$  satisfying

$$\lambda_{a,b}^x(\omega_s^t \neq \eta_s^t) \leq \gamma_\phi^t \mathbb{1}_{\mathcal{L}(x,0)}(s, t),$$

for all  $t \in \mathbb{N}$  and  $s \in \mathbb{L}$ .

Here  $\mathbb{1}_{\mathcal{L}(x,0)}$  denotes the characteristic function of the light cone  $\mathcal{L}(x,0)$  and  $\omega, \eta$  are random variables in  $\Omega$  distributed according to the marginals of  $\lambda_{a,b}^x$ .

*Proof.* We follow the same lines as in Theorem 8.3.1. The coupling  $\lambda_{a,b}^x$  is a PCA on  $\Omega \times \Omega$ , defined by the probability transition function  $\bar{\phi}$  given in (8.3.1). The PCA starts with the coupling  $\lambda \in \mathcal{J}_T(\mu_a^x, \mu_b^x)$  with support in  $\Delta_{a,b}^x := \{(\omega, \eta) \in X \times \Omega : \omega_s = \eta_s \forall s \neq x, \omega_x = a \text{ and } \eta_x = b\}$ , defined as follows. For each  $\Lambda \in \mathbb{L}$ , and  $a, b \in S^\Lambda$  we have

$$\lambda([a] \times [b]) = \begin{cases} (\mu_a[a] - \mu_b[b])^+ + (\mu_b[b] - \mu_a[a])^+ + \min(\mu_a[a], \mu_b[b]) & \text{if } [a] \times [b] \cap \Delta_{a,b}^x \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that  $\lambda([a] \times \Omega) = \mu_a[a]$  and  $\lambda(\Omega \times [b]) = \mu_b[b]$ . Let us denote by  $\lambda^\tau \in \mathcal{M}_1(\Omega \times \Omega)$  the marginal of  $\lambda_{a,b}^x$  on the coordinates in  $(\mathbb{L} \times \mathbb{L}) \times \{\tau\}$ , by  $\lambda_1^\tau \in \mathcal{M}(\Omega)$  the marginal of  $\lambda^\tau$  on the coordinates of the first copy of  $\mathbb{L}$ ,

and similarly for  $\lambda_2^\tau \in \mathcal{M}(\Omega)$ . Now, for each  $\Lambda \in \mathbb{L}$  and  $\tau \in \mathbb{N}$  we have

$$\begin{aligned} \lambda_1^{\tau+1}[a] &= \lambda^{\tau+1}([a] \times X) = \sum_{b \in S^\Lambda} \sum_{c, d \in S^{\Lambda+U}} \lambda^\tau([c] \times [d]) \prod_{x \in \Lambda} \bar{\phi}(a_x, c_{x+U}; b_x, d_{x+U}) \\ &= \sum_{c, d \in S^{\Lambda+U}} \lambda^\tau([c] \times [d]) \prod_{x \in \Lambda} \left( \sum_{b_x \in A} \bar{\phi}(a_x, c_{x+U}; b_x, d_{x+U}) \right) \\ &= \sum_{c \in S^{\Lambda+U}} \lambda_1^\tau[c] \prod_{x \in \Lambda} \phi(a_x, c_{x+U}) = \mathcal{P}_\Phi \lambda_1^\tau[a]. \end{aligned}$$

An analogous computation gives us  $\lambda_2^{\tau+1}[a] = \mathcal{P}_\Phi \lambda_2^\tau[a]$  for each  $b \in S^{\Lambda+U}$ . Since  $\lambda^0 = \lambda \in \mathcal{J}_T(\mu_a^x, \mu_b^x)$ , by induction on  $\tau$  we readily deduce that  $\lambda^\tau \in \mathcal{J}_T(\mathcal{P}_\Phi^\tau \mu, \mathcal{P}_\Phi^\tau \nu)$ .

At time  $\tau = 0$ ,  $s \neq x$  is equivalent to  $(s, 0) \notin \mathcal{L}_{(x,0)}$  and for such  $s$  we have  $\lambda^0(\omega_s^0 \neq \eta_s^0) = 0$ . Let us assume that for  $\tau \in \mathbb{N}$  and  $(s, \tau) \notin \mathcal{L}_{(x,0)}$ , we have  $\lambda^\tau(\omega_s^\tau \neq \eta_s^\tau) = 0$ . Then at time  $\tau + 1$ ,

$$\begin{aligned} \lambda_{a,b}^x(\omega_s^{\tau+1} \neq \eta_s^{\tau+1}) &= \sum_{a, b \in S^{\{s\}}, a \neq b} \lambda^{\tau+1}([a] \times [b]) \\ &= \sum_{c, d \in S^{U+s}} \sum_{a \neq b} \bar{\phi}(a, c; b, d) \lambda^\tau([c] \times [d]) \\ &= \sum_{c, d \in S^{U+s}} \sum_{c \neq d} \frac{\lambda^\tau([c] \times [d])}{\|\phi_c - \phi_d\|_{TV}} \sum_{a \in A} (\phi_c - \phi_d)^+(a) \sum_{b \neq a} (\phi_d - \phi_c)^+(b) \\ &\leq \sum_{c, d \in S^{U+s}} \sum_{c \neq d} \lambda^\tau([c] \times [d]) \|\phi_c - \phi_d\|_{TV}. \end{aligned}$$

If  $(s, \tau + 1) \notin \mathcal{L}_{(x,0)}$  then necessarily  $(U + s) \times \{\tau\} \cap \mathcal{L}_{(x,0)} = \emptyset$  and therefore  $\lambda^\tau([c] \times [d]) = 0$  for each  $c, d \in S^{U+s}$  such that  $c \neq d$ . Otherwise, if  $((U + s) \times \{\tau\}) \cap \mathcal{L}_{(x,0)} \neq \emptyset$ , following the computations in the proof of Theorem 8.3.1, for each  $c, d \in S^{U+x}$  with  $c \neq d$  we have

$$\lambda^\tau([c] \times [d]) \|\phi_c - \phi_d\|_{TV} \leq \lambda^\tau([c] \times [d]) \sum_{s \in U} \max\{\|\phi_e - \phi_{e'}\|_{TV} : e_\zeta = e'_\zeta, \zeta \neq s\}.$$

Let  $\gamma_s := \max\{\|\phi_e - \phi_{e'}\|_{TV} : e_\zeta = e'_\zeta, \zeta \neq s\}$ , then

$$\begin{aligned} \sum_{c, d \in S^{U+s}} \sum_{c \neq d} \lambda^\tau([c] \times [d]) \|\phi_c - \phi_d\|_{TV} &\leq \sum_{c, d \in S^{U+s}} \sum_{c \neq d} \lambda^\tau([c] \times [d]) \sum_{c_{\zeta+s} \neq d_{\zeta+s}} \gamma_\zeta \\ &\leq \sum_{\zeta \in U} \gamma_\zeta \sum_{c, d \in S^{U+s}: c_{\zeta+s} \neq d_{\zeta+s}} \lambda^\tau([c] \times [d]) \\ &= \sum_{\zeta \in U} \gamma_\zeta \lambda^\tau(\omega_{\zeta+s}^\tau \neq \eta_{\zeta+s}^\tau) \\ &\leq \gamma_\phi \sup_{(\zeta, \tau) \in \mathcal{L}_{(x,0)}} \lambda_{a,b}^x(\omega_\zeta^\tau \neq \eta_\zeta^\tau). \end{aligned}$$

In this way we obtain

$$\lambda_{a,b}^x(\omega_s^{\tau+1} \neq \eta_s^{\tau+1}) \leq \gamma_\phi \sup_{(\zeta, \tau) \in \mathcal{L}_{(x,0)}} \lambda_{a,b}^x(\omega_\zeta^\tau \neq \eta_\zeta^\tau) \mathbb{1}_{\mathcal{L}_{(x,0)}}(s, t).$$

Since this inequality holds for all an arbitrary  $s \in \mathbb{L}$ , it follows by induction in  $\tau$  that

$$\lambda_{a,b}^x(\omega_s^t \neq \eta_s^t) \leq \gamma_\phi^t \sup_{(s,0) \in \mathcal{L}_{(x,0)}} \lambda_{a,b}^x(\omega_s^0 \neq \eta_s^0) \mathbb{1}_{\mathcal{L}_{(x,0)}}(s, t) \leq \gamma_\phi^t \mathbb{1}_{\mathcal{L}_{(x,0)}}(s, t)$$

for each  $t \in \mathbb{N}$  and  $s \in \mathbb{L}$ .  $\square$

We can now proceed to the proof of Theorem 8.4.1.

**Proof of Theorem 8.4.1.** Let us fix  $K \in \mathcal{L}_B$ . For each  $\epsilon > 0$  let  $\Lambda \in \mathbb{L}$  and  $\mathcal{T} \in \mathbb{N}$  be such that in the funnel  $F = F_\Lambda^{\mathcal{T}}$  we have

$$\|\Delta K\|_2^2 - \sum_{(x,t) \in F} |\Delta K|^2 < \epsilon.$$

With this we define the local observable  $\mathcal{K} : X \rightarrow \mathbb{R}$  given by

$$\mathcal{K}(\omega) = \max\{K(\eta) : \eta \in [\omega_F]\}.$$

From now on we will consider the local observable  $\mathcal{K}$  approaching the original space-time observable  $K$ . The proof follows the classical argument which consists of decomposing  $\mathcal{K} - \mathbb{E}(\mathcal{K})$  as a martingale sum (see [21] for instance). Let  $n : F \rightarrow \{1, 2, \dots, |F|\}$  be an ordering of  $F$ . For each  $1 \leq k \leq |F|$ , let  $D_k \mathcal{K} : S^k \rightarrow \mathbb{R}$  be such that

$$D_k \mathcal{K}(\omega_{n \leq n(k)}) := \mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) - \mathbb{E}(\mathcal{K} | \omega_{n < n(k)}),$$

where by convention  $\mathbb{E}(\mathcal{K} | \omega_{n < n(1)}) \equiv \mathbb{E}(\mathcal{K})$  and  $\mathbb{E}(\mathcal{K} | \omega_{n \leq n(|F|)}) \equiv \mathcal{K}(\omega_F) = \mathcal{K}(\omega)$ . With this,

$$\mathcal{K}(\omega) - \mathbb{E}(\mathcal{K}) = \sum_{k=1}^{|F|} D_k \mathcal{K}(\omega_{n \leq n(k)}). \quad (8.4.1)$$

In order to prove the theorem, we will use Azuma-Hoeffding (see [53, Page. 68]), which ensures that

$$\mathbb{E}(e^{\mathcal{K} - \mathbb{E}(\mathcal{K})}) \leq e^{\frac{1}{2} \sum_{k=1}^{|F|} \|D_k \mathcal{K}\|_\infty^2}. \quad (8.4.2)$$

In order to bound  $D_k \mathcal{K}$  we note that

$$\begin{aligned} D_k \mathcal{K}(\omega_{n \leq n(k)}) &= \mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) - \mathbb{E}(\mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) | \omega_{n < n(k)}) \\ &\leq \max_{\omega_{n(k)}} \mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) - \mathbb{E}\left(\min_{\omega_{n(k)}} \mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) | \omega_{n < n(k)}\right) \\ &= \max_{\omega_{n(k)}} \mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) - \min_{x_{n(k)}} \mathbb{E}(\mathcal{K} | \omega_{n \leq n(k)}) =: \Delta_k \mathcal{K}(\omega_{n < n(k)}). \end{aligned}$$

It is now convenient to express the max-min difference  $\Delta_k \mathcal{K}(\omega_{n < n(k)})$  in terms of a coupling. For this, we have to consider the spatiotemporal decomposition of the coordinates in  $F$  and we have to impose a monotonicity condition for the ordering  $n$ . For each  $1 \leq k \leq |F|$ , let  $n(k) = (x(k), t(k)) \in \mathbb{L} \times \mathbb{N}$ . We require from  $n$  to be time-increasing, *i.e.*,  $t(k) \geq t(k-1)$  for all  $2 \leq k \leq |F|$ . For each  $a \in S$  and  $1 \leq k \leq |F|$ , let  $\mu_a \in \mathcal{M}(\Omega)$  be such that

$$\mu_a[\omega_E] := \mathcal{P}_\Phi^{t(k)} \mu \left( [\omega_E \mid [\omega_{n < n(k), t=t(k)} a]] \right)$$

for each  $E \in \mathbb{L}$ . To be more clear,  $\mu_a$  is obtained at time  $t(k)$  by the action of the PCA on  $\mu$ , and then conditioning the resulting distribution on the cylinder  $[\omega_{n < n(k)} a]$ . For  $a \neq b \in S$  and  $1 \leq k \leq |F|$  let  $\lambda_{a,b}^k \in \mathcal{J}_T(\mathbb{P}_{\mu_a}, \mathbb{P}_{\mu_b})$  be the coupling between  $\mu_a$  and  $\mu_b \in \mathcal{M}(\Omega)$  described in Lemma 8.4.1. The coupling  $\lambda_{a,b}^k$  is a PCA on  $\Omega \times \Omega$  whose marginals are precisely realizations of the original PCA starting at  $\mu_a$  and  $\mu_b$  respectively, hence,

$$D_k \mathcal{K}(\omega_{n < n(k)}) = \max_{a,b \in S} \int_\Omega (\mathcal{K}(\omega_{n < n(k)} a u) - \mathcal{K}(\omega_{n < n(k)} b w)) d \lambda_{a,b}^k(u, v). \quad (8.4.3)$$

We follow now the standard trick consisting in reducing the difference

$$\mathcal{K}(\omega_{n < n(k)} a u) - \mathcal{K}(\omega_{n < n(k)} b w)$$

to a sum of differences involving a change in a single site. For each  $v, w \in S^{\{n > n(k)\}}$  we have

$$\begin{aligned} & \mathcal{K}(\omega_{n < n(k)} a u) - \mathcal{K}(\omega_{n < n(k)} b w) \\ &= \sum_{j=k+1}^{|F|} \mathcal{K}(\omega_{n < n(k)} a u_{n \leq n(j)} w_{n > n(j)}) - \mathcal{K}(\omega_{n < n(k)} b u_{n < n(j)} w_{n \geq n(j)}). \end{aligned}$$

Here  $u_{n \leq n(j)} w_{n > n(j)}$  denotes the configuration in  $S^{\{n > n(k)\}}$  which coincides with  $u$  at the first  $j - k$  coordinates and with  $w$  at the rest of the coordinates. Similarly for  $u_{n < n(j)} w_{n \geq n(j)}$ . From this we have

$$\left| \mathcal{K}(\omega_{n < n(k)} a u) - \mathcal{K}(\omega_{n < n(k)} b w) \right| \leq \Delta_{n(k)} \mathcal{K} + \sum_{j=k+1}^{|F|} \delta(v_{n(j)}, w_{n(j)}) \Delta_{n(j)} \mathcal{K},$$

where  $\delta(\cdot, \cdot)$  denotes the Kronecker's delta. Then, the integral in (8.4.3) can be bounded as follows,

$$\left| D_k \mathcal{K}(\omega_{n < n(k)}) \right| \leq \Delta_{n(k)} \mathcal{K} + \sum_{j=k+1}^{|F|} \max_{a, b \in S} \lambda_{a, b}^k (u_{n(j)} \neq v_{n(j)}) \Delta_{n(j)} \mathcal{K}.$$

According to Lemma 8.4.1, the coupling  $\lambda_{a, b}^k$  satisfies

$$\begin{aligned} \lambda_{a, b}^k (u_{n(j)} \neq v_{n(j)}) &\leq \gamma_\phi^{t(j) - t(k)} \sup_{x \in \mathbb{L}} \lambda_{a, b}(\omega_x^{t(k)} \neq \eta_x^{t(j)}), \text{ if } n(j) \in \mathcal{L}_{n(k)}, \\ \lambda_{a, b}^k (u_{n(j)} \neq v_{n(j)}) &= 0, \text{ otherwise.} \end{aligned}$$

Here  $\mathcal{L}_{n(k)}$  is the light cone emerging from  $n(k) \in F$ , as defined in (8.2.4). With this,

$$\begin{aligned} \left| D_k \mathcal{K}(\omega_{n < n(k)}) \right| &\leq \Delta_{n(k)} \mathcal{K} + \sum_{n(j) \in \mathcal{L}_{n(k)}} \gamma_\phi^{t(j) - t(k)} \Delta_{n(j)} \mathcal{K} \\ &= \sum_{j=k}^{|F|} \gamma_\phi^{t(j) - t(k)} \mathbb{1}_{\mathcal{L}_{n(k)}}(n(j)) \Delta_{n(j)} \mathcal{K}, \end{aligned} \quad (8.4.4)$$

where  $\mathbb{1}_{\mathcal{L}_{n(k)}}$  denotes the characteristic function of the light cone  $\mathcal{L}_{n(k)}$ . Now, the summation in the right-hand side of this inequality can be seen as the  $k$ -th coordinate of a vector  $\zeta \in \mathbb{R}^{\{1, 2, \dots, |F|\}}$  which is obtained as a matrix product

$$\zeta := \Gamma_\phi \Delta \mathcal{K}, \quad (8.4.5)$$

where  $\Gamma_\phi : \{1, 2, \dots, |F|\} \times \{1, 2, \dots, |F|\} \rightarrow \mathbb{R}$  is the upper triangular matrix

$$\Gamma_\phi(k, \ell) = \begin{cases} \gamma_\phi^{t(\ell) - t(k)} \mathbb{1}_{\mathcal{L}_{n(k)}}(n(\ell)) & \text{if } \ell \geq k, \\ 0 & \text{otherwise} \end{cases}$$

and  $\Delta \mathcal{K} \in \mathbb{R}^{\{1, 2, \dots, |F|\}}$  has coordinates  $(\Delta \mathcal{K})_k := \Delta_{n(k)} \mathcal{K}$ . Hence, we can rewrite (8.4.4) as  $\|D_k \mathcal{K}\|_\infty \leq \Gamma_\phi \Delta \mathcal{K}$ , and from this we obtain

$$\frac{1}{2} \sum_{k=1}^{|F|} \|D_k \mathcal{K}\|_\infty^2 \leq \frac{1}{2} \|\Gamma_\phi \Delta \mathcal{K}\|_2^2 \leq \frac{\|\Gamma_\phi\|_2^2}{2} \sum_{k=1}^{|F|} |\Delta_{n(k)} \mathcal{K}|^2, \quad (8.4.6)$$

where  $\|\Gamma_\phi\|_2$  is the  $\ell_2$  induced norm of the matrix  $\Gamma_\phi$ . It is well-known that for a matrix  $M$ ,  $\|M\|_2^2$  coincides with the largest eigenvalue of the product  $(M M^\dagger)$ . On the other hand, the maximal eigenvalue of a non-negative matrix is bounded by the maximum of its row sums, thus,

$$\begin{aligned} \|\Gamma_\phi\|_2^2 &\leq \max_{1 \leq k \leq |F|} \sum_{\ell=1}^{|F|} \sum_{j=1}^{|F|} \gamma_\phi^{2t(j)-t(k)-t(\ell)} \mathbb{1}_{\mathcal{L}_{n(k)}}(n(j)) \mathbb{1}_{\mathcal{L}_{n(\ell)}}(n(j)) \\ &= \max_{1 \leq k \leq |F|} \sum_{\mathcal{L}_{n(k)} \cap \mathcal{L}_{n(\ell)} \neq \emptyset} \sum_{n(j) \in \mathcal{L}_{n(k)} \cap \mathcal{L}_{n(\ell)}} \gamma_\phi^{2t(j)-t(k)-t(\ell)}. \end{aligned}$$

Taking into account that the largest light cones emerge at time  $t = 0$ , and that all other cones emerging at points in the funnel  $F$  have to intersect a light cone emerging at  $t = 0$ , then

$$\|\Gamma_\phi\|_2^2 \leq \max_{t(k)=0} \sum_{\ell=1}^{|F|} \sum_{n(j) \in \mathcal{L}_{n(\ell)} \cap \mathcal{L}_{n(k)}} \gamma_\phi^{2t(j)-t(\ell)}.$$

At this point it is convenient to point out some easily verifiable facts about the light cones and funnels:

- a) If  $(\zeta, \tau) \in \mathcal{L}_{(x,t)}$  then  $\tau \geq t$  and  $(x, t) \in F_\zeta^\tau$ .
- b) For each  $(x, t) \in \mathbb{L} \times \mathbb{N}$  we have

$$\mathcal{L}_{(x,t)} = \cup_{(\zeta,\tau) \in \mathcal{L}_{(x,t)}} \mathcal{L}_{(\zeta,\tau)}.$$

- c) For each  $(\zeta, \tau) \in \mathbb{L} \times \mathbb{N}$  and  $t \leq \tau$  we have

$$|F_\zeta^\tau \cap \mathbb{L} \times \{t\}| \leq (\tau - t) \text{diam}(U) + 1.$$

- d) For each  $(x, t) \in \mathbb{L} \times \mathbb{N}$  and  $\tau \geq t$  we have

$$|\mathcal{L}_{(x,t)} \cap \mathbb{L} \times \{\tau\}| \leq (\tau - t) \text{diam}(U) + 1.$$

Taking all this into account, if  $t(k) = 0$  then

$$\begin{aligned} \sum_{\ell=1}^{|F|} \sum_{n(j) \in \mathcal{L}_{n(\ell)} \cap \mathcal{L}_{n(k)}} \gamma_\phi^{2t(j)-t(\ell)} &\leq \sum_{n(j) \in \mathcal{L}_{n(k)}} \sum_{(x,t) \in F_\zeta^\tau} \gamma_\phi^{2(t(j)-t(\ell))} \gamma_\phi^{t(\ell)} \\ &\leq \sum_{\tau=0}^T \gamma_\phi^\tau (\tau \text{diam}(U) + 1) \sum_{t=0}^{\tau} \gamma_\phi^{2(\tau-t)} ((\tau - t) \text{diam}(U) + 1) \\ &\leq \sum_{\tau=0}^{\infty} \gamma_\phi^\tau (\tau \text{diam}(U) + 1) \sum_{t=0}^{\infty} \gamma_\phi^{2t} (t \text{diam}(U) + 1) \\ &\leq \frac{\text{diam}(U)^2}{(1 - \gamma_\phi^2)^2} \sum_{\tau=0}^{\infty} \gamma_\phi^\tau (\tau \text{diam}(U) + 1) \leq \frac{\text{diam}(U)^2}{(1 - \gamma_\phi)^2 (1 - \gamma_\phi^2)^2}. \end{aligned}$$

This upper bound does not depend on  $k$ , as long as  $t(k) = 0$ . Hence

$$\|\Gamma_\phi\|_2^2 \leq \frac{\text{diam}(U)^2}{(1 - \gamma_\phi)^2 (1 - \gamma_\phi^2)^2}$$

With this we can rewrite (8.4.6) as

$$\frac{1}{2} \sum_{k=1}^{|F|} \|D_k \mathcal{K}\|_\infty^2 \leq \frac{\text{diam}(U)^2 \sum_{k=1}^{|F|} |\Delta_{n(k)} K|^2}{2(1-\gamma_\phi)^2(1-\gamma_\phi^2)^2}.$$

Applying (8.4.2) we obtain

$$\mathbb{E} \left( e^{\mathcal{K} - \mathbb{E}(\mathcal{K})} \right) \leq e^{\frac{\text{diam}(U)^2}{2(1-\gamma_\phi)^2(1-\gamma_\phi^2)^2} \sum_{k=1}^{|F|} |\Delta_{n(k)} \mathcal{K}|^2} \leq e^{\frac{\text{diam}(U)^2}{2(1-\gamma_\phi)^2(1-\gamma_\phi^2)^2} \|\Delta K\|_2^2}.$$

So far we have proved that for each observable  $K \in \mathcal{L}_B$  and for each  $\epsilon > 0$ , there exists a local function  $\mathcal{K} : X \rightarrow \mathbb{R}$  such that

$$0 \leq \|\Delta K\|_2^2 - \|\Delta \mathcal{K}\|_2^2 < \epsilon$$

and

$$\mathbb{E} \left( e^{\mathcal{K} - \mathbb{E}(\mathcal{K})} \right) \leq e^{\frac{\text{diam}(U)^2}{2(1-\gamma_\phi)^2(1-\gamma_\phi^2)^2} \|\Delta K\|_2^2}.$$

Now, if  $\omega_F = \eta_F$  then

$$|K(\omega) - K(\eta)| \leq \sum_{(x,t) \notin F} \Delta_{(x,t)} K = \|\Delta K\|_2^2 - \sum_{(x,t) \in F} |\Delta_{(x,t)} K|^2 < \epsilon.$$

Hence  $\|K - \mathcal{K}\|_\infty < \epsilon$ , which implies

$$K - \mathbb{E}(K) \leq \mathcal{K} - \mathbb{E}(\mathcal{K}) + 2\epsilon.$$

From this it follows

$$\mathbb{E} \left( e^{K - \mathbb{E}(K)} \right) \leq e^{\frac{\text{diam}(U)^2}{2(1-\gamma_\phi)^2(1-\gamma_\phi^2)^2} \|\Delta K\|_2^2} e^{2\epsilon}$$

for each  $\epsilon > 0$ . The theorem follows from this, by taking

$$C = \frac{\text{diam}(U)^2}{2(1-\gamma_\phi)^2(1-\gamma_\phi^2)^2}.$$

□



# Chapter 9

## Perspectives

We list some questions we would like to think about in the near future.

### 9.1 Concentration inequalities for marginal distributions of the joint distribution of the PCA

In the high-noise regime, we could prove that an ergodic PCA satisfies a spatio-temporal Gaussian concentration bound for all initial distributions. This is due to the weakly dependent interaction nature of the system in this range of “noise”. This regime only considers the variations of transition probabilities by changing a spin value in the past and, for this reason, it does not correspond exactly to the Dobrushin uniqueness regime (see 2.3) for the spatio-temporal measure. In this sense, even if, the high-noise condition does not ensure the Gibbsianity of the unique invariant shift-invariant marginal measure, one may ask if it satisfies a spatial Gaussian concentration bound. Thus, we formulate the following open question.

**CONJECTURE 2.** *Assume that  $\mu^*$  is the unique invariant shift-invariant measure for the PCA and  $\mathbb{P}_\mu$  satisfies 8.2.12 for all initial distribution  $\mu \in \mathcal{M}(\Omega)$ . Then there exists  $D = D(\mu^*) > 0$  such that for all functions  $K : \Omega \rightarrow \mathbb{R}$  in  $\mathcal{L}$ , we have*

$$\mathbb{E}_{\mu^*} [\exp (K - \mathbb{E}_{\mu^*} [K])] \leq \exp (D \|\underline{\delta}(K)\|_2^2).$$

Other situations might be interesting to study and correspond to the non-ergodic cases of PCAs where there exists a unique non-attractive probability measure and where there exist several invariant measures. For those cases, we expect the following results.

**CONJECTURE 3.** *Assume that there exist several invariant measures for the PCA, then the PCA does not satisfy a spatio-temporal Gaussian concentration bound*

**CONJECTURE 4.** *Assume that there exist a unique non-attractive measure for the PCA, then the PCA does not satisfy a spatio-temporal Gaussian concentration bound*

### 9.2 Coupled map lattices

This is about multidimensional lattices of weakly coupled piecewise expanding interval maps. Given  $I \subset \mathbb{R}$  compact interval, we will consider a dis-



crete time dynamics on the phase space  $\Omega = I^{\mathbb{Z}^d}$ . It is composed of independent chaotic actions on each component and of some weak interaction between the components that does not destroy the chaotic character of the whole system. More precisely, let  $\tau : I \rightarrow I$  be the single site dynamics. We assume this application to be  $\mathcal{C}^2$  from  $I$  to  $I$  with singularities at  $\zeta_1, \dots, \zeta_{N-1} \in (0, 1)$  in the sense that  $\tau$  is monotone and  $\mathcal{C}^2$  on each component of  $I \setminus \{\zeta_0 = 0, \zeta_1, \dots, \zeta_{N-1}, \zeta_N = 1\}$ . Moreover, we assume that  $\tau', \tau''$  are bounded and  $\inf |\tau'| > 2$  (uniformly expanding). We define the unperturbed dynamics  $T_0 : \Omega \rightarrow \Omega$  by  $[T_0(x)]_p = \tau(x_p)$ .

Now we define the perturbed dynamics introducing a coupling  $\Phi_\epsilon : \Omega \rightarrow \Omega$  of the form  $\Phi_\epsilon(x) = x + A_\epsilon(x)$ . We say that such coupling has a range  $r$  and strength  $\epsilon$  if for all  $k, p, q \in \mathbb{Z}^d$

$$|(A_\epsilon)_p|_\infty \leq 2\epsilon, \quad |(DA_\epsilon)_{qp}|_\infty \leq 2\epsilon, \quad |\partial_k(DA_\epsilon)_{qp}|_\infty \leq 2\epsilon, \quad (9.2.1)$$

and  $\partial_p \Phi_{\epsilon,q} = 0$  whenever  $|p - q| > r$ . We want to study the dynamics  $T_\epsilon : \Omega \rightarrow \Omega$  defined by

$$T_\epsilon := \Phi_\epsilon \circ T_0. \quad (9.2.2)$$

We are interested in the concentration properties of the process  $(x, T_\epsilon x, \dots)$  with  $x$  distributed according to the unique  $T_\epsilon$ -invariant measure  $\mu_\epsilon$ . In [45, 3] the authors proved the following Ruelle-Perron-Frobenius theorem which quantifies the exponential convergence of the iterates of the transfer operator on the space of Lipschitz functions  $\mathcal{L}$ .

**Theorem 9.2.1** ([45, 3]).

For all  $r \in \mathbb{N}$ , there exists  $\epsilon_1(r) > 0$  such that, if  $\Phi_\epsilon$  is a coupling with range  $r$  and strength  $0 \leq \epsilon \leq \epsilon_1(r)$ , Then we have:

- there exists un unique pair  $(h_\epsilon, \mu_\epsilon)$  such that  $h_\epsilon$  is a bounded variation function and is strictly positive,  $\int h_\epsilon dm = 1$ . Moreover  $P_\epsilon h_\epsilon = h_\epsilon$  and  $P_\epsilon^* \mu_\epsilon = \mu_\epsilon$ .
- Let  $\Lambda \Subset \mathbb{Z}^d$ . There exists a constant  $C(\Lambda)$  and  $\gamma \in (0, 1)$ ,  $C_{(9.2.1)} > 0$  such that for every function  $f \in \mathcal{L}$  depending on a finite number of coordinates in  $\Lambda$ ,

$$\|P_\epsilon^n f - \mu_\epsilon(f)\|_\infty \leq C_{(9.2.1)} C(\Lambda) \gamma^n \|f\|_{\mathcal{L}}, \quad \forall n \in \mathbb{N}. \quad (9.2.3)$$

They also proved a central and local limit theorem for this system.

**Theorem 9.2.2** ([3]).

For all  $r \in \mathbb{N}$ , there exists  $\epsilon_1(r) > 0$  satisfying the following properties. If  $\Phi_\epsilon$  is a coupling with range  $r$  and strength  $0 \leq \epsilon \leq \epsilon_1(r)$  and  $\mu_\epsilon$  is the unique invariant measure. Let  $f \in \mathcal{L}$  depending on a finite number of coordinates in  $\Lambda$  with  $\int f d\mu_\epsilon = 0$ .

Then there exists  $\sigma^2 \geq 0$  such that  $\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f \circ T_\epsilon^k$  converges in distribution to  $\mathcal{N}(0, \sigma^2)$  with respect to the measure  $\mu_\epsilon$ . Moreover,  $\sigma^2 = 0$  if and only if there exists a measurable function  $u : \Omega \rightarrow \mathbb{R}$  such that  $f = u - u \circ T_\epsilon$   $\mu_\epsilon$ -almost everywhere.

If we also assume that, whenever  $u : \Omega \rightarrow \mathbb{R}$  is measurable and  $\lambda \in \mathbb{R}^*$ , the function  $f - u + u \circ T_\epsilon \text{ mod } \lambda$  is not  $\mu_\epsilon$ -almost everywhere constant - we say that  $f$  is aperiodic. In particular, the variance  $\sigma^2$  in the central limit theorem is nonzero. Then, for any compact interval  $I \subset \mathbb{R}$ ,

$$\sigma \sqrt{2\pi n} \mu_\epsilon \{x : S_n f(x) \in I\} \rightarrow |I|$$

where  $|I|$  is the length of the interval  $I$ .

We may ask for which condition on the interaction, the unique Lebesgue measure satisfies a Gaussian concentration bound. Thus, we present the following conjecture.

**CONJECTURE 5.** *If  $\Phi_\epsilon$  is a coupling with range  $r$  and strength  $0 \leq \epsilon \leq \epsilon_1(r)$  and  $\mu_\epsilon$  is the unique invariant measure. Then, there exists  $D > 0$  such that for all functions  $K : \Omega^t \rightarrow \infty$  in  $\mathcal{L}_B$ , one has*

$$\mathbb{E}_{\mathbb{P}_{\mu_\epsilon}} \left[ e^{K(x^{[0:t-1]}) - \mathbb{E}_{\mathbb{P}_{\mu_\epsilon}} [K(x^{[0:t-1]})]} \right] \leq e^{D \|\Delta(K)\|_2^2} \quad (9.2.4)$$

with  $D > 0$  a positive constant depending only on  $\epsilon$  and  $d$ .

### 9.3 Countable Markov shifts

Here, we present the study of the thermodynamic formalism for countable Markov chains introduced by O. Sarig (see [75]). More precisely, it concerns the statistical properties of the equilibrium states associated to the shift map on sequences on a countable alphabet and a sufficiently regular potential  $\Phi : \Omega \rightarrow \mathbb{R}$ . He proved that the potential satisfies a condition of positive recurrence is a necessary and sufficient condition for a Ruelle-Perron-Frobenius theorem to hold. He also showed that under big image property condition of the topological Markov shift i.e.

$$\exists b_1, \dots, b_n \in S \forall a \in S \exists i [a, b_i] \neq \emptyset.$$

the convergence of the iterate of the transfer operator is uniform exponential. In addition, in her Ph.D. thesis Veronique Maume (see [66]) proved a similar result by using Birkhoff cones technics with the following condition on  $\Phi$ :

$$\exists k_1, n_1 \in \mathbb{N} \quad \text{such that } \forall k > k_1, \exists \rho_k < 1 \quad \text{such that } \forall n > n_1, P_\Phi^k \mathbf{1}_n \leq \rho_k$$

where  $P_\Phi$  is the Ruelle-Perron-Frobenius operator associated to  $\Phi$ . Such convergence echoes the Gaussian concentration inequalities conditions in the case of sub-shift of finite type studied by Jean-René Chazottes and Sébastien Gouezel in [17]. We conjecture the following result.

**CONJECTURE 6.** *Under these conditions, there exists  $D > 0$  such that for any separately Lipschitz observables  $K : \Omega^n \rightarrow \mathbb{R}$ , we have*

$$\mathbb{E}_\mu \left[ e^{K(x, Tx, \dots, T^{n-1}x) - \mathbb{E}_\mu [K(x, Tx, \dots, T^{n-1}x)]} \right] < \exp \left( D \sum_{j=0}^{n-1} \text{Lip}_j(K)^2 \right).$$

For countable Markov chains J. Dedecker and S. Gouëzel proved in [26] that an irreducible aperiodic Markov chain is geometrically ergodic if and only

if it satisfies GCB for separately bounded functions. One may ask if there exists a similar criteria of GCB for countable Markov shifts or if there exist conditions on the Markov chains (non-uniqueness of the measure, null recurrent measure) which prevents GCB to hold.

## 9.4 Relationship between GCB and complete analyticity

In Chapter 6, we could prove that complete analyticity of the potential implies that the associated unique measure satisfies a Gaussian concentration bound. We think that there exists a more direct proof of such a result that we were not able to prove it. We expect the following.

**CONJECTURE 7.** *Let  $\Phi \in \mathcal{B}_R$ . The Gibbs measures  $\gamma_\Lambda^\Phi$  satisfy the weak mixing condition for all  $\Lambda \in \mathbb{Z}^d$  then  $\mu_\Phi$  satisfies a Gaussian concentration bound.*

We also think that there might exist an equivalence between complete analyticity and Gaussian concentration bound for all specifications.

**CONJECTURE 8.** *Let  $\Phi \in \mathcal{B}_R$ . The specification  $\gamma_\Lambda^\Phi$  satisfies  $\text{GCB}(D)$  with finite constant  $D = \sup_{\Lambda \in \mathbb{Z}^d} D(\mu, \Lambda) < +\infty$  implies that Complete analyticity holds for  $\Phi$ .*

For potentials  $\Phi \in \mathcal{B}_T$ , one may ask if there exist equivalence theorems or modestly implications between the notions of weak mixing, complete analyticity, restricted complete analyticity, logarithmic Sobolev inequalities and Gaussian concentration bounds.

# Bibliography

- [1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion operators*. Grundlehren der mathematischen Wissenschaften, Vol. 348. Springer, Jan. 2014.
- [2] V. Baladi. *Positive Transfer Operators and Decay of Correlations*. Advanced series in nonlinear dynamics. World Scientific Publishing Company, 2000.
- [3] J.-B. Bardet, S. Gouëzel, and G. Keller. Limit theorems for coupled interval maps. *Stochastics and Dynamics*, 07, 03 2007.
- [4] A. Basuev. Ising model in half-space: A series of phase transitions in low magnetic fields. *Theoretical and Mathematical Physics*, 153:1539–1574, 11 2007.
- [5] J. L. Beck and K. M. Zuev. Rare event simulation, 2015.
- [6] S. Berghout, R. Fernandez, and E. Verbitskiy. On the relation between gibbs and  $g$ -measures. *Ergodic Theory and Dynamical Systems*, 39(12):3224–3249, 2019.
- [7] R. Bissacot, E. O. Endo, A. C. D. van Enter, and A. Le Ny. Entropic repulsion and lack of the  $g$ -measure property for dyson models. *Communications in Mathematical Physics*, 363(3):767–788, Sep 2018.
- [8] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.
- [9] C. Borgs. Translation symmetry breaking in four-dimensional lattice gauge theories. *Comm. Math. Phys.*, 96(2):251–284, 1984.
- [10] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [11] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.
- [12] P. Bullen. *Dictionary of inequalities*. Monographs and Research Notes in Mathematics. CRC Press, Boca Raton, FL, second edition, 2015.
- [13] J.-R. Chazottes, P. Collet, C. Külske, and F. Redig. Concentration inequalities for random fields via coupling. *Probab. Theory Related Fields*, 137(1-2):201–225, 2007.

- [14] J.-R. Chazottes, P. Collet, and F. Redig. On concentration inequalities and their applications for Gibbs measures in lattice systems. *Journal of Statistical Physics*, 169(3):504–546, Nov 2017.
- [15] J.-R. Chazottes, P. Collet, and B. Schmitt. Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. *Nonlinearity*, 18(5):2341–2364, 2005.
- [16] J.-R. Chazottes, S. Gallo, and D. Takahashi. Optimal gaussian concentration bounds for stochastic chains of unbounded memory, 2020.
- [17] J.-R. Chazottes and S. Gouëzel. Optimal concentration inequalities for dynamical systems. *Comm. Math. Phys.*, 316(3):843–889, 2012.
- [18] J.-R. Chazottes and C. Maldonado. Concentration bounds for entropy estimation of one-dimensional Gibbs measures. *Nonlinearity*, 24(8):2371–2381, 2011.
- [19] J.-R. Chazottes, J. Moles, F. Redig, and E. Ugalde. Gaussian concentration and uniqueness of equilibrium states in lattice systems, 2020.
- [20] J.-R. Chazottes, J. Moles, and E. Ugalde. Gaussian concentration bound for potentials satisfying Walters condition with subexponential continuity rates. *Nonlinearity*, 33(3):1094–1117, Jan 2020.
- [21] J. R. Chazottes and F. Redig. Concentration inequalities for Markov processes via coupling, 2008.
- [22] J.-R. Chazottes and E. Ugalde. Entropy estimation and fluctuations of hitting and recurrence times for Gibbsian sources. *Discrete Contin. Dyn. Syst. Ser. B*, 5(3):565–586, 2005.
- [23] P. Collet and A. Galves. Chains of infinite order, chains with memory of variable length, and maps of the interval. *J. Stat. Phys.*, 149(1):73–85, 2012.
- [24] P. Collet, S. Martinez, and B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probability Theory and Related Fields*, 123(3):301–322, 2002.
- [25] L. Coquille, A. C. D. van Enter, A. Le Ny, and W. M. Ruszel. Absence of dobrushin states for 2d long-range ising models. *Journal of Statistical Physics*, 172(5):1210–1222, Jun 2018.
- [26] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic markov chains. *Electron. Commun. Probab.*, 20:12 pp., 2015.
- [27] M. Denker, C. Grillenberger, and K. Sigmund. *Ergodic Theory on Compact Spaces*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2006.
- [28] R. L. Dobrushin and S. B. Shlosman. *Completely Analytical Gibbs Fields*, pages 371–403. Birkhäuser Boston, Boston, MA, 1985.

- [29] R. L. Dobrushin and S. B. Shlosman. Completely analytical interactions: Constructive description. *Journal of Statistical Physics*, 46(5):983–1014, 1987.
- [30] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains. Operation research and financial engineering*. Springer, 2018.
- [31] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, Cambridge, 2009.
- [32] R. Dudley. *Real Analysis and Probability*. CRC Press, 2018.
- [33] R. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Classics in Mathematics. Springer, 2006.
- [34] R. Fernandez, S. Gallo, and G. Maillard. Regular g-measures are not always gibbsian. *Electronic Communications in Probability*, 16, 06 2011.
- [35] R. Fernández and A. Galves. Markov approximations of chains of infinite order. *Bulletin of the Brazilian Mathematical Society*, 33(3):295–306, 2002.
- [36] R. Fernández and G. Maillard. Chains with complete connections and one-dimensional Gibbs measures. *Electronic Journal of Probability*, 9:145–176, Jan. 2004. 31 pages.
- [37] S. Friedli and Y. Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, 2017.
- [38] S. Gallo and D. Y. Takahashi. Attractive regular stochastic chains: perfect simulation and phase transition. *Ergodic Theory Dynam. Systems*, 34(5):1567–1586, 2014.
- [39] H. Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter studies in mathematics. W. de Gruyter, 1988.
- [40] S. Goldstein, R. Kuik, J. Lebowitz, and C. Maes. From pca’s to equilibrium systems and back. *Communications in Mathematical Physics*, 125:71–79, 01 1989.
- [41] H. Hennion and L. Herve. *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2003.
- [42] Y. Holovatch. *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory*. Number vol. 3 in Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory. World Scientific, 2004.
- [43] D. Ioffe. Exact large deviation bounds up to  $T_c$  for the Ising model in two dimensions. *Probability Theory and Related Fields*, 102(3):313–330, 1995.
- [44] G. Keller. *Equilibrium states in ergodic theory*, volume 42 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.

- [45] G. Keller and C. Liverani. Uniqueness of the srb measure for piecewise expanding weakly coupled map lattices in any dimension. *Communications in Mathematical Physics*, 262:33–50, 02 2006.
- [46] A. Kondah, V. Maume, and B. Schmitt. Vitesse de convergence vers l'état d'équilibre pour des dynamiques Markoviennes non Höldériennes. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6):675–695, 1997.
- [47] A. Kondah, V. Maume, and B. Schmitt. Vitesse de convergence vers l'état d'équilibre pour des dynamiques markoviennes non höldériennes. *Annales de l'I.H.P. Probabilités et statistiques*, 33(6):675–695, 1997.
- [48] A. Kontorovich and M. Raginsky. Concentration of measure without independence: a unified approach via the martingale method. *ArXiv*, abs/1602.00721, 2016.
- [49] J. Kotze. Introduction to monte carlo methods for an ising model of a ferromagnet, 2008.
- [50] C. Külske. Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures. *Communications in Mathematical Physics*, 239(1):29–51, 2003.
- [51] D. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 2015.
- [52] J. L. Lebowitz, C. Maes, and E. R. Speer. Statistical mechanics of probabilistic cellular automata. *Journal of Statistical Physics*, 59(1):117–170, 1990.
- [53] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [54] F. Ledrappier. Principe variationnel et systèmes dynamiques symboliques. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 30(3):185–202, 1974.
- [55] T. Lindvall. *Lectures on the Coupling Method*. Dover Books on Mathematics. Dover Publications, 2012.
- [56] P.-Y. Louis. Ergodicity of pca: Equivalence between spatial and temporal mixing conditions. *Electronic Communications in Probability*, 9(0):119–131, 2004.
- [57] C. Maes and S. B. Shlosman. Ergodicity of probabilistic cellular automata: a constructive criterion. *Comm. Math. Phys.*, 135(2):233–251, 1991.
- [58] J. Mairesse and I. Marcovici. Around probabilistic cellular automata. *Theoretical Computer Science*, 559:42 – 72, 2014. Non-uniform Cellular Automata.
- [59] I. Marcovici, M. Sablik, and S. Taati. Ergodicity of some classes of cellular automata subject to noise. *Electronic Journal of Probability*, 24(0), 2019.

- [60] F. Martinelli and E. Olivieri. Approach to equilibrium of glauber dynamics in the one phase region. i. the attractive case. *Comm. Math. Phys.*, 161(3):447–486, 1994.
- [61] F. Martinelli, E. Olivieri, and R. H. Schonmann. For 2-d lattice spin systems weak mixing implies strong mixing. *Communications in Mathematical Physics*, 165(1):33–47, 1994.
- [62] K. Marton. Bounding  $\bar{d}$ -distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 04 1996.
- [63] K. Marton. A measure concentration inequality for contracting markov chains. *Geometric & Functional Analysis GAFA*, 6(3):556–571, 1996.
- [64] K. Marton. Measure concentration for a class of random processes. *Probability Theory and Related Fields*, 110(3):427–439, 1998.
- [65] K. Marton and P. C. Shields. The positive-divergence and blowing-up properties. *Israel Journal of Mathematics*, 86(1):331–348, 1994.
- [66] V. Maume-Deschamps. *Propriétés de mélange pour des systèmes dynamiques markoviens*. PhD thesis, Université de Bourgogne, 1998.
- [67] B. McCoy and T. Wu. *The Two-Dimensional Ising Model: Second Edition*. Dover Publications, 2014.
- [68] A. Ny. Gibbs measures for long-range ising models. *arXiv: Mathematical Physics*, 2019.
- [69] S. Ott. Weak mixing and analyticity of the pressure in the ising model. *Communications in Mathematical Physics*, 377(1):675–696, Oct 2019.
- [70] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.
- [71] M. Pollicott. Rates of mixing for potentials of summable variation. *Transactions of the American Mathematical Society*, 352:843–853, 2000.
- [72] D. Ruelle. *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*. Cambridge Mathematical Library. Cambridge University Press, 2004.
- [73] L. Rueschendorf and T. Sei. On optimal stationary couplings between stationary processes. *Electron. J. Probab.*, 17:20 pp., 2012.
- [74] P.-M. Samson. Concentration of measure inequalities for Markov chains and  $\phi$ -mixing processes. *Ann. Probab.*, 28(1):416–461, 01 2000.
- [75] O. M. Sarig. Thermodynamic formalism for countable markov shifts. *Ergodic Theory and Dynamical Systems*, 19(6):1565–1593, 1999.
- [76] R. H. Schonmann. Second order large deviation estimates for ferromagnetic systems in the phase coexistence region. *Comm. Math. Phys.*, 112(3):409–422, 1987.
- [77] R. H. Schonmann and S. B. Shlosman. Complete analyticity for 2d ising completed. *Comm. Math. Phys.*, 170(2):453–482, 1995.



- [78] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.
- [79] J. E. Steif. Convergence to equilibrium and space time bernoullicity for spin systems in the  $M < \epsilon$  case. *Ergodic Theory and Dynamical Systems*, 11(3):547–575, 1991.
- [80] D. W. Stroock and B. Zegarliński. The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition. *Comm. Math. Phys.*, 144(2):303–323, 1992.
- [81] D. W. Stroock and B. Zegarliński. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.*, 149(1):175–193, 1992.
- [82] A. C. D. van Enter, R. Fernández, and A. D. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory. *Journal of Statistical Physics*, 72(5-6):879–1167, Sep 1993.
- [83] P. Walters. Ruelle’s operator theorem and  $g$ -measures. *Trans. Amer. Math. Soc.*, 214:375–387, 1975.
- [84] P. Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.*, 236:121–153, 1978.
- [85] K. M. Zuev. *Subset Simulation Method for Rare Event Estimation: An Introduction*, pages 1–25. Springer Berlin Heidelberg, Berlin, Heidelberg, 2021.

