Universidade Federal da Bahia - UFBA
Instituto de Matemática e Estatística - IME
Programa de Pós-Graduação em Matemática - PGMAT
Tese de Doutorado

On Weak and Strong Large Deviation Principles for the Empirical Measure of Random Walks

Joedson de Jesus Santana

Salvador-Bahia

February 21, 2025

Ficha catalográfica elaborada pela Biblioteca Universitária de
Ciências e Tecnologias Prof. Omar Catunda, SIBI – UFBA.

S232 Santana, Joedson de Jesus

Sobre princípios de grandes desvios fracos e fortes para a medida empírica de passeios aleatórios / Joedson de Jesus Santana. – Salvador, 2025.

68 f.

Orientador: Prof. Dr. Dirk Erhard

Coorientador: Prof. Dr. Tertuliano Franco Santos Franco

Tese (Doutorado) – Universidade Federal da Bahia. Instituto de Matemática e Estatística, 2025.

1. Princípio de Grandes Desvios. 2. Passeio aleatório. 3. Medida Empírica. 4. Topologia de Mukherjee e Varadhan. I. Erhard, Dirk. II. Franco, Tertuliano Franco Santos. III. Universidade Federal da Bahia. IV. Título.

CDU: 519.222

On Weak and Strong Large Deviation Principles for the Empirical Measure of Random Walks

Joedson de Jesus Santana

Tese de Doutorado apresentada ao Colegiado do Programa de Doutorado em Matemática em Associação UFBA/UFAL, como requisito parcial para obtenção do Título de Doutor em Matemática.

Orientador: Prof. Dr. Dirk Erhard.

Salvador-Bahia

February 21, 2025

"Sobre Princípios de Grandes Desvios Fracos e Fortes para a Medida Empírica de Passeios Aleatórios."

JOEDSON DE JESUS SANTANA

Tese apresentada ao Colegiado do Curso de Pós-graduação em Matemática da Universidade Federal da Bahia, como requisito parcial para obtenção do Título de Doutor em Matemática.

Banca Examinadora

Documento assinado digitalmente
DIRK ERHARD
Data: 01/03/2025 09:43:21-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Dirk Erhard (Orientador - UFBA)

Otávio de Macedo Menezes

Prof. Dr. Otavio de Macedo Menezes (UFBA)

Documento assinado digitalmente
TERTULIANO FRANCO SANTOS FRANCO
Data: 25/02/2025 10:35:09-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Tertuliano Franco Santos Franco (UFBA)

Conrado Freitas Paulo da Costa

Prof. Dr. Conrado Freitas Paulo da Costa (DURHAM UNIVERSITY)

Documento assinado digitalmente
ADRIANA NEUMANN DE OLIVEIRA
Data: 25/02/2025 09:54:58-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Adriana Neuman de Oliveira (UFRGS)

Agradecimentos

A Deus, pelo dom da vida e por me permitir chegar até aqui.

Às minhas mães Luciene e Ildenir.

Ao meu pai Gerson.

Aos meus irmãos e irmãs Lucimeire, Gerson, Emerson, Cátia, Diego e Carine.

Ao meu orientador, Dirk Erhard, pela orientação, amizade, paciência, oportunidade, confiança e apoio.

Ao meu coorientador Tertuliano.

Aos meus professores, em especial Eudson, Eduardo, Monteiro, Rosane, Elisângela, Liliane, Ricardo, Vanessa, Manuela, Jerome, Joílson, Jaime, Samuel, Henrique, Cristina e Armando

Aos amigos de sempre e aos novos, em especial, Ualas, Djavan, Douglas, Rose, Renata, Diva, Wellington, Iasmim, Carlos, Gleide, Maic, Mariza, Vinícius, Ícaro, Benner, Sávio, Elivan, Janara, João, Rafael (Coordena), Renan, Diegão, Leandro, Paulo, Cleide, Enio, Glaene e Claudio, Leydiane, Rafael, Diego, Carlos, Iago, Eldon, Eduardo, Tio Nivaldo e Cris, Carina, Elaine, Juan, Armando e Sheyla, Adan, Milton.

À família Palheta que me acolheu: Nara, Henderson, Hellen, Hellene e Henrique.

Aos colegas professores que estão sempre dispostos a me ajudar: Átila, Rosemir, Ricardo, Wagner, Manoel, Edvaldo, Luciana, Robert e Ronildo.

À minha namorada Thaís, por todo incentivo e apoio.

Aos meus alunos.

Por fim, agradeço à CAPES pelo apoio financeiro.

“ Digam o que disserem o mal do século é a solidão.
Cada um de nós imerso em sua própria arrogância
esperando por um pouco de afeição.”
Renato Russo

Abstract

This work is divided into two chapters. In the first chapter, we provide an introduction to the theory of large deviations and prove a weak Large Deviation Principle (LDP) for the empirical measure of the random walk with certain rates. To achieve this, we use the Parabolic Anderson Model (PAM) and the Gärtner-Ellis Theorem. In the second chapter we show that the empirical measure of certain continuous time random walks satisfies a strong large deviation principle with respect to a topology introduced in [21] by Mukherjee and Varadhan. This topology is natural in models which exhibit an invariance with respect to spatial translations. Our result applies in particular to the case of simple random walk and complements the results obtained in [21] in which the large deviation principle has been established for the empirical measure of Brownian motion.

Keywords: Large Deviation Principle, random walk, empirical measure, topology of Mukherjee and Varadhan.

Resumo

Este trabalho está dividido em dois capítulos. No primeiro capítulo, apresentamos uma introdução à teoria dos grandes desvios e demonstramos um Princípio de Grandes Desvios (LDP) fraco para a medida empírica do passeio aleatório com certas taxas. Para isso, utilizamos o Modelo Parabólico de Anderson (PAM) e o Teorema de Gärtner-Ellis. No segundo capítulo, demonstramos que a medida empírica de certos passeios aleatórios em tempo contínuo satisfaz um princípio de grandes desvios forte com respeito a uma topologia introduzida por Mukherjee e Varadhan em [21]. Essa topologia é natural em modelos que exibem invariância em relação a translações espaciais. Nosso resultado aplica-se, em particular, ao caso do passeio aleatório simples e complementa os resultados obtidos em [21], nos quais o princípio dos grandes desvios foi estabelecido para a medida empírica do movimento Browniano.

Palavras-chave: Princípio de Grandes Desvios, passeio aleatório, medida empírica, topologia de Mukherjee e Varadhan.

Contents

Chapter 1

Introduction and a Weak Large Deviation Principle

1.1 Introduction

In this section we give a short introduction to the theory of large deviations. Given the extensive body of literature now available, our treatment of the general theory will be concise. We refer the reader to [6, 7] for excellent sources offering an in-depth study of the general theory. Our primary focus will be on the large deviation theory for empirical measures of random walks. In particular, we will show that for a broad class of random walks, their empirical measures satisfy what is known as a weak large deviation principle. While this result is likely familiar to experts in the field, we were not able to find a proof of this in the literature. The study of large deviation principles of empirical measures of random walks is interesting and relevant, since as we will illustrate with two examples, there are many models in probability theory and statistical mechanics that are defined in terms of empirical measures. Since in statistical mechanics one often looks at exponential changes of measures, it is important to understand the asymptotics of unlikely events, and this is what large deviation theory is about. In this chapter we will focus on what is known as weak large deviation principle, and in the next chapter we will focus on a new topology under which one can prove a strong large deviation principle. The latter is useful in contexts in which one wants to understand finer properties of the model at hand, opposed to the crude understanding of the partition function or similar objects.

1.1.1 A brief review on large deviation theory

In an introductory course on probability theory, two fundamental results are typically studied: the law of large numbers and the central limit theorem. The former characterizes the average behaviour of a large number of independent repetitions of a given experiment, while the latter describes the fluctuations around this average behaviour. Both results focus on the *typical* outcomes. In contrast, large deviation theory investigates the occurrence of *atypical* events. For example, consider a sequence of independent coin tosses. If the coin is fair, the law of large numbers asserts that, after n tosses, approximately $n/2$ will result in head. The central limit theorem refines this by describing the typical deviations from $n/2$: it states that the number of heads is typically given by $n/2 + \sqrt{n}\mathcal{N}(0, \frac{1}{2}) + o(\sqrt{n})$. In large deviation theory, as the name suggests, the goal is to characterize the probabilities of extreme events, such as:

$$\{\text{the number of heads is at least } an\} \qquad \text{and} \qquad \{\text{the number of heads is at most } bn\}\,, \tag{1.1.1}$$

where $a > \frac{1}{2}$ and $b < \frac{1}{2}$. To study this specific example, consider a sequence $X_1, X_2, \dots$ of i.i.d Bernoulli random variables with parameter $\frac{1}{2}$ and let $S_n = \sum_{i=1}^n X_i$. For $a \in (\frac{1}{2}, 1]$ it is known that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a)\,, \tag{1.1.2}$$

where

$$I(a) = \sup_{t\in\mathbb{R}}[ta - \log \varphi(t)] = \log 2 + a \log a + (1-a)\log(1-a)\,, \tag{1.1.3}$$

and φ is the moment generating function of X_1, i.e., $\varphi(t) = \mathbb{E}[e^{tX_1}]$. The function I is referred to as rate function and it often happens to be the case that it is given by the solution to a variational problem. It however happens less often that one can explicitly solve this problem.

Before we proceed with formal definitions, let us briefly discuss why studying large deviations is useful. In statistical mechanics one often begins with a reference process, such as a simple random walk $S = (S_n)_{n\in\mathbb{N}}$ or a Brownian motion $B = (B_t)_{t\geq 0}$. Given a Hamiltonian H, which is a function of the process, one studies the sequence of probability measures $\mathbb{P}_H$ defined by

$$\mathbb{P}_H(S_{[0,n]} \in A) = \frac{\mathbb{E}\Big[\exp(H(S_{[0,n]}))\mathbb{1}_{\{S_{[0,n]}\in A\}}\Big]}{Z_{H,n}}\,, \tag{1.1.4}$$

where $Z_{H,n}$ is a normalization constant ensuring that $\mathbb{P}_H$ is a probability measure. The first example for a Hamiltonian H is given by

$$H(S_{[0,n]}) = \beta \sum_{i=0}^{n} \xi(S_i)\,, \tag{1.1.5}$$

for a field $\{\xi(x) : x \in \mathbb{Z}^d\}$ of i.i.d. random variables and a constant $\beta > 0$. This Hamiltonian corresponds to a polymer measure and is the discrete analogue of the parabolic Anderson model which will be introduced later on. The second example involves a Hamiltonian H acting on continuous functions $B : [0,\infty) \to \mathbb{R}$ via

$$H(B_t) = \frac{1}{t}\int_0^t ds \int_0^t du \, \frac{1}{|B_s - B_u|}\,. \tag{1.1.6}$$

This choice in three dimensions leads to the so-called Polaron model, see for instance [22, 18]. The above Hamiltonian of course only has a chance to be finite for continuous functions B without any self-intersection which is the case for Brownian motion in dimension three, the dimension of interest, and it is indeed possible to show, see [21], that the division by t is the correct thing to do if one wants to observe non-trivial behaviors. This will be more clear later on when we connect the above model with the study of empirical measures.

The study of both models naturally involves large deviations. In the first example, the measure $\mathbb{P}_H$ favors trajectories that visit regions where ξ is large. While such behaviour may be probabilistically costly under the random walk measure, it could become typical under $\mathbb{P}_H$. Analysing $\mathbb{P}_H$ requires balancing the probabilistic cost of the random walk travelling far to reach these regions against the potential gains of doing so.

A similar situation arises in the Polaron model, which is related to the second example above. The polaron problem describes the motion of a charged particle (e.g., an electron) in a crystal, where lattice polarization influences the particle's effective behavior. We refer to the lectures of Feynman [13] for a more detailed description of the problem. Coming back to the Hamiltonian in (1.1.6), we observe that in three dimension Brownian motion is neighborhood transient, meaning it does not necessarily revisit specific regions. However, the Hamiltonian favors trajectories that return to previously visited points. Indeed, if the trajectory approaches previously visited points, the value of H increases, thereby enhancing the expectation on the right-hand side of (1.1.4). The behavior of thus balances the probabilistic cost of deviating from typical Brownian motion against the energetic rewards of revisiting spatial locations.

We now come to more precise definitions. To that end we denote by $\mathcal{X}$ a metric space, that along the text we might equip with additional properties. In particular we can talk about open and closed sets.

Definition 1.1. *Given a metric space $\mathcal{X}$. A function $I : \mathcal{X} \to [0,\infty]$ is called a **rate function** if*

(R$_1$) $I \not\equiv \infty\,,$

(R_2) *I is lower semi-continuous.*

If in addition to (R_2) *one has*

(R_3) *I has compact level sets,*

then I is called a good rate function.

Remark 1.2. *Some authors require the rate function in the above definition always to be good, as in [7] for example. In this monograph, however, we follow the conventions of [6].*

Next we define the notion of the Large Deviation Principle (LDP). Let $(\gamma_n)_n$ be a sequence of positive real numbers such that $\gamma_n \to \infty$ as $n \to \infty$.

Definition 1.3. *A sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ on $\mathcal{X}$ satisfies a Large Deviation Principle (LDP) with rate function I and rate (γ_n) if the following conditions hold:*

(L_1) *I is a rate function*

(L_2) *For every open set $G \subseteq \mathcal{X}$ we have that*

$$\liminf_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(G) \geq - \inf_{x\in G} I(x)\,. \tag{1.1.7}$$

(L_3) *For every closed set $F \in \mathcal{X}$ we have that*

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(F) \leq - \inf_{x\in F} I(x)\,. \tag{1.1.8}$$

Remark 1.4. *The above definition closely resembles the definition of weak convergence of probability measures. The key difference is that here we have convergence on an exponential scale.*

Remark 1.5. *The above definition works without any major modification in continuous time. If in that case the rate (γ_t) is given by $\gamma_t = t$, then we say that the rate of the LDP is t.*

The upper bound in (L_3) for closed sets is generally derived from the upper bound for compact sets, combined with an argument of exponential tightness. We now remind the reader of the definition of exponential tightness.

Definition 1.6. *A sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ on $\mathcal{X}$ is said to be exponentially tight if for any $\alpha > 0$ there is a compact set K_ε such that*

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(\mathcal{X} \setminus K_\varepsilon) < -\alpha\,. \tag{1.1.9}$$

Remark 1.7. *Some authors define the above notion of exponential tightness only for the specific case in which the rate is given by $\gamma_n = n$. However, in cases in which one expects a rate different to that one, the above notion is more useful.*

It often happens that the sequence at hand is not exponentially tight and the upper bound fails for closed sets. In this case we also speak of a weak large deviation principle. More precisely, one has the following definition.

Definition 1.8. *A sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ satisfies a weak Large Deviation Principle (weak LDP) with rate function I and scale (γ_n) if* (L_1) *and* (L_2) *in Definition 1.3 are satisfied and if additionally one has that*

(L_3^*) *For every compact set $K \in \mathcal{X}$ we have that*

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(K) \le -\inf_{x\in K} I(x). \tag{1.1.10}$$

In the next section, we will focus on LDPs for empirical measures of random walks.

1.1.2 Large Deviation Principle for the Empirical Measure

Let $(X_t)_{t\ge 0}$ be a continuous time Markov chain on $\mathbb{Z}^d$ with distribution $\mathbb{P}$ and generator $\mathcal{L}$ acting on bounded functions $f : \mathbb{Z}^d \to \mathbb{R}$ taking the form

$$(\mathcal{L}f)(x) = \sum_{y\in\mathbb{Z}^d} a_{x,y}[f(y) - f(x)]\,. \tag{1.1.11}$$

We denote the domain of $\mathcal{L}$ by $\mathcal{D}$ and we assume that the rates $(a_{x,y})_{x,y\in\mathbb{Z}^d}$ satisfy the following assumptions.

(A1) The process $X = (X_t)_{t\ge 0}$ is a Feller process.

(A2) $\sup_{x\in\mathbb{Z}^d} \sum_{y\in\mathbb{Z}^d} a_{x,y} < \infty$.

(A3) For all $x, y \in \mathbb{Z}^d$ there is $n \in \mathbb{N}$ and a sequence of points $x = x_0, x_1, \dots, x_n = y$ such that for all $i \in \{0, 1, \dots, n-1\}$ we have that $a_{x_i,x_{i+1}} > 0$.

(A4) There exists a positive function $(u_t)_{t\ge 0}$ satisfying $\lim_{t\to\infty} \frac{1}{t} \log u_t = 0$ such that

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\Big(\sup_{0\le s\le t} |X_s| \ge u_t \Big) = -\infty.$$

(A5) The rates are symmetric in the sense that for all $x, y \in \mathbb{Z}^d$ one has that $a_{x,y} = a_{y,x}$.

Remark 1.9. *We remark that the above conditions are in particular satisfied for the symmetric simple random walk and for any symmetric random walk with finite range if it is irreducible. In this case it suffices to choose $u_t = t\log t$. If the random walk does not have a finite range, then extra work is needed to verify the above conditions.*

We are interested in the study of the *local times*, defined by

$$\ell_t(x) = \int_0^t \mathbb{1}_{\{X_s=x\}}\, ds, \quad x \in \mathbb{Z}^d,\ t > 0\,, \tag{1.1.12}$$

which measures the amount of time the process spends at x until time t. We are interested in the large deviations of the local times defined above and to that end it turns out to be more convenient to look at the normalized local times, namely,

$$\frac{\ell_t(x)}{t} = \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=x\}} ds =: L_t(x)\,, \tag{1.1.13}$$

so that $L_t = \{L_t(x) : x \in \mathbb{Z}^d\}$ becomes a probability measure and is referred to as the *empirical measure*. We denote the space of probability measures on $\mathbb{Z}^d$ by $\mathcal{M}_1(\mathbb{Z}^d)$.

To formulate the main result of this section we define a rate function I via

$$I(\mu) = \left\|(-\mathcal{L})^{\frac{1}{2}}\sqrt{\mu}\right\|^2 = \frac{1}{2}\sum_{x,y\in\mathbb{Z}^d} a_{x,y}(\sqrt{\mu}(x) - \sqrt{\mu}(y))^2\,, \tag{1.1.14}$$

where the norm appearing above is on the space $\ell^2(\mathbb{Z}^d)$, and the second equality follows from a straightforward computation using the symmetry of the rates.

Under the above assumptions, we will prove the following result.

Theorem 1.10. *Let $X = (X_t)_{t\geq 0}$ be a Markov chain satisfying assumptions* (A1) – (A5)*. The family of probabilities measures $(L_t)_{t\geq 0}$ satisfies a weak large deviations principle with respect to the usual weak convergence of probability measures in $\mathcal{M}_1(\mathbb{Z}^d)$ with rate function I defined in* (1.1.14)*, and rate t.*

The lack of the upper bound for closed sets in the above result follows from a lack of exponential tightness. Indeed, by Prokhorov's theorem a subset $\mathcal{A} \subseteq \mathcal{M}_1(\mathbb{Z}^d)$ is pre-compact if and only if it is tight. The latter means that for any $\varepsilon > 0$ there is a compact set $K \subseteq \mathbb{Z}^d$ such that

$$\sup_{\mu\in\mathcal{A}} \mu(K^c) < \varepsilon\,. \tag{1.1.15}$$

It is not hard to convince oneself that given $\varepsilon \in (0,1)$ there is no compact set $K \subseteq \mathbb{Z}^d$

such that the event

$$\Big\{L_t(K^c) \geq \varepsilon\Big\} \tag{1.1.16}$$

has a probability that tends to zero exponentially in t. Indeed, the probability that the Markov chains eventually escapes any given compact set tends to one. Before we come to the proof of Theorem 1.10 in Section 1.2 we will present one application of the (LDP) for local times. We will moreover, lay the theoretical groundwork for it by presenting a version of the Gärtner-Ellis theorem that will be particularly useful for us.

The Parabolic Anderson Model (PAM)

For a motivation of the study of large deviations of the empirical measure of the symmetric simple random walk, let us consider the following Cauchy problem for the heat equation:

$$\begin{cases} \partial_t u(t,z) = \Delta^d u(t,z) + \xi(z)u(t,z), & \text{for } (t,z) \in (0,\infty) \times \mathbb{Z}^d\,, \\ u(0,z) = \delta_0(z) \;\; \text{for } z \in \mathbb{Z}^d. \end{cases} \tag{1.1.17}$$

Here, $\xi = \{\xi(z)\,:z \in \mathbb{Z}^d\}$ is a field of real-valued of i.i.d random variables. The operator Δ^d is the discrete Laplacian defined via

$$\Delta^d f(z) = \sum_{y\in\mathbb{Z}^d: y\sim z} [f(y) - f(z)] \;\; \text{for } z \in \mathbb{Z}^d,\; f \in \ell^2(\mathbb{Z}^d) \tag{1.1.18}$$

where $y \sim z$ means that y and z are nearest neighbours. It is known from [14] that the above equation has a unique non-negative solution under mild assumptions on the potential. This problem is known in the literature as the Parabolic Anderson Model (PAM). Its non-negative solution has a natural interpretation in terms of branching random walks: consider a random walk starting at the origin and jumping at rate $2d$. At site x at time t, the walk branches into two independent random walks at rate $\xi^+(x)$ and dies at rate $\xi^-(x)$. Here, ξ^+ and ξ^- represent the positive and negative parts of ξ, respectively. In this context, $u(t,x)$ denotes the expected number of branching random walks at site x at time t. Before exploring the relationship between the PAM and the empirical measure of a simple random walk, we will first provide some basic background on the model.

The Feynman-Kac Formula

One of the features that make the PAM amenable to a detailed analysis is the fact that its solution has an explicit probabilistic representation in terms of the Feynman-Kac formula, which reads

$$u(t,z) = \mathbb{E}_0\left[\exp\left\{\int_0^t \xi(X(s))\mathrm{d}s\right\}\delta_z(X(t))\right],\ (t,z)\in[0,\infty)\times\mathbb{Z}^d. \tag{1.1.19}$$

Here, $(X(s))_{s\in[0,\infty)}$ is a continuous-time random walk on $\mathbb{Z}^d$ with generator Δ^d starting at $z\in\mathbb{Z}^d$ under $\mathbb{E}_z$. From the above formula, we see that the most significant contributions come from random walk paths that balance the probabilistic cost of reaching regions where the potential is high with the gains from such a strategy.

Summing the above over all $z\in\mathbb{Z}^d$ we obtain the total mass of the solution denoted by $U^\xi(t)$, which can be written as follows:

$$U^\xi(t) = \mathbb{E}_0\left[\exp\left\{\int_0^t \xi(X(s))\mathrm{d}s\right\}\right],\ t\in[0,\infty). \tag{1.1.20}$$

Local Times and Moments

We now provide the connection between the PAM and the local times $(\ell_t)_{t\geq 0}$ defined in Section 1.1.2

First note that a straightforward computation shows that

$$\int_0^t \xi(X_s)\mathrm{d}s = \sum_{z\in\mathbb{Z}^d}\xi(z)\ell_t(z). \tag{1.1.21}$$

Hence, denoting by $\langle\ \rangle$ the expectation with respect to ξ, i.e., given a function F^ξ which depends in a measurable way on ξ we write $\langle F^\xi\rangle$ to denote its expectation with respect to the field ξ, and recalling that the random potential ξ is i.i.d., we can calculate

$$\left\langle \exp\left\{\int_0^t \xi(X_s)\mathrm{d}s\right\}\right\rangle = \left\langle \prod_{z\in\mathbb{Z}^d} e^{\xi(z)\ell_t(z)}\right\rangle = \prod_{z\in\mathbb{Z}^d}\left\langle e^{\xi(0)\ell_t(z)}\right\rangle. \tag{1.1.22}$$

Here, in the last equality we used the fact that the expectation is with respect to ξ and not to ℓ_t. Defining the logarithmic moment generating function H via

$$H(t) = \log\left\langle e^{t\xi(0)}\right\rangle, \tag{1.1.23}$$

we see that the right hand side of (1.1.22) equals

$$\exp\left\{\sum_{z\in\mathbb{Z}^d} H(\ell_t(z))\right\}. \tag{1.1.24}$$

Assuming that $H(t)$ is finite for all positive t, Fubini's theorem yields that

$$\left\langle U^{\xi}(t)\right\rangle = \mathbb{E}_0\left[\exp\left\{\sum_{z\in\mathbb{Z}^d} H(\ell_t(z))\right\}\right]. \tag{1.1.25}$$

The above calculations establish the relation between local times and the PAM. Next, we will analyze the asymptotic behaviour for the term in (1.1.25). To that end we introduce one condition about the random environment ξ that was fundamental in [14].

Assumption 1.11. *There exists* $0 \le \rho < \infty$ *such that*

$$\lim_{t\to\infty} \frac{H(ct) - cH(t)}{t} = \rho c \log c \;\forall c \in (0,1) \tag{1.1.26}$$

and the convergence is uniform on $[0,1]$.

Before we continue we remind the reader of Varadhan's lemma which in its simplest form can be stated as follows.

Lemma 1.12 (Varadhan's Lemma)**.** *Let* $\mathcal{X}$ *be a Polish space and* (μ_n) *be a family of probability measures in* $\in \mathcal{M}_1(X)$ *satisfying the LDP on* $\mathcal{X}$ *with rate* γ_n *and rate function* I*. Let* $F : \mathcal{X} \to \mathbb{R}$ *be a continuous function that is bounded from above. Then*

$$\lim_{n\to\infty} \frac{1}{\gamma_n} \log \int_{\mathcal{X}} e^{\gamma_n F(x)} \mu_n(\mathrm{d}x) = \sup_{x\in\mathcal{X}} [F(x) - I(x)]. \tag{1.1.27}$$

Before we pick up our computations from (1.1.25) we define

$$J(\mu) = -\sum_{x\in\mathbb{Z}^d} \mu(x)\log\mu(x). \tag{1.1.28}$$

Since the probability measure L_t was defined as ℓ_t/t, and therefore $\sum_{z\in\mathbb{Z}^d} L_t(z) = 1$, we can rewrite (1.1.25) as

$$\left\langle U^{\xi}(t)\right\rangle = e^{H(t)}\mathbb{E}_0 \exp\left\{t\sum_{z\in\mathbb{Z}^d} \frac{H(L_t(z)t) - L_t(z)H(t)}{t}\right\}. \tag{1.1.29}$$

By Assumption 1.11 above, the expression

$$\sum_{z\in\mathbb{Z}^d} \frac{H(L_t(z)t) - L_t(z)H(t)}{t} \tag{1.1.30}$$

should becomes uniformly close to $\sum_{z\in\mathbb{Z}^d} \rho L_t(z)\log L_t(z)$ as $t \to \infty$. Therefore, we obtain the following

$$\langle U^\xi(t)\rangle = e^{H(t)+o(t)}\mathbb{E}_0 \exp\{-t\rho J(L_t)\}\,. \tag{1.1.31}$$

Now, supposing that we have a LDP for $(L_t)_{t\geq 0}$ with rate function I and rate t, Varadhan's Lemma 1.12 yields that

$$\lim_{t\to\infty}\frac{1}{t}\log \mathbb{E}_0 \exp\{-t\rho J(L_t)\} = -\inf_{\mu\in\mathcal{M}_1(\mathbb{Z}^d)}[I(\mu)+\rho J(\mu)]. \tag{1.1.32}$$

Together with (1.1.31) this is enough to determine the asymptotic behaviour of $\mathbb{E}_0 \exp\{-t\rho J(L_t)\}$.

Remark 1.13. *The attentive reader might have noticed that our computations were not completely rigorous. Indeed, first of all Assumption 1.11 does not guarantee that* (1.1.30) *converges, since the sum therein is an infinite sum. Second,* $(L_t)_{t\geq 0}$ *does not satisfy a strong large deviation principle. The above result however is still correct and a rigorous proof uses a compactification argument that we will not repeat here. We refer the reader to [14] for more details. The above example however, still illustrates the usefulness of a large deviation principle for the empirical measure.*

1.1.3 Gärtner-Ellis-Theorem

In this section we introduce one of the key results in the theory of large deviations and an adaptation of it that is suitable for our purposes. Before we dive into the details we will introduce some necessary definitions and observations.

Definition 1.14. *Let $\mathcal{X}$ be a Hausdorff (real) topological vector space and denote by $\mathcal{X}^*$ the space of all continuous linear functionals on $\mathcal{X}$. Let $(Z_t)_{t\geq 0}$ be a family of random variables taking values in $\mathcal{X}$, and for each $t\geq 0$ we denote by $\mu_t\in M_1(\mathcal{X})$ the distribution of Z_t. The logarithmic moment generating function $\Lambda_{\mu_t}: X^*\to(-\infty,\infty]$ is defined via*

$$\Lambda_{\mu_t}(V) = \log \mathbb{E}\left[e^{\langle V, Z_t\rangle}\right] = \log\int_{\mathcal{X}} e^{V(x)}\mu_t(\mathrm{d}x)\,, \tag{1.1.33}$$

where for $x\in\mathcal{X}$ and $V\in\mathcal{X}^*$, $\langle V,x\rangle$ denotes the value of $V(x)\in\mathbb{R}$. We define

$$\bar\Lambda(V) := \limsup_{t\to\infty}\frac{1}{\gamma_t}\Lambda_{\mu_t}\Big(\gamma_t V\Big) \tag{1.1.34}$$

and write $\Lambda(V)$ *whenever the limit exists. The Fenchel-Legendre transform of* $\Lambda(V)$ *is defined as*

$$\Lambda^*(x) = \sup_{V \in \mathcal{X}^*} \{\langle V, x\rangle - \Lambda(V)\}. \tag{1.1.35}$$

For the local times $\ell_t(z)$ and the respective empirical measure $L_t(z) = \dfrac{\ell_t(z)}{t}$ and given a bounded function V defined on $\mathbb{Z}^d$, the above becomes

$$\Lambda_{L_t}(V) = \log \mathbb{E}\left[e^{\langle V, L_t\rangle}\right] = \log \mathbb{E}\left[e^{\left(\sum_{z\in\mathbb{Z}^d} V(z) L_t(z)\right)}\right], \tag{1.1.36}$$

and in this way

$$\bar{\Lambda}(V) = \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[e^{\sum_{z\in\mathbb{Z}^d} V(z)L_t(z)t}\right] = \limsup_{t\to\infty} \frac{1}{t}\Lambda_{L_t}(tV). \tag{1.1.37}$$

We note that the expression $\Lambda_{L_t}(tV)$ equals U^{ξ^t}, where for each $t \geq 0$, the potential ξ^t is given by $\xi^t(x) = tV(x)$.

Definition 1.15. *Let* X, Y *be Banach spaces and* $U \subset X$ *be an open set. We say that* $f : U \to Y$ *is Gâteaux differentiable at* $x \in U$ *in the direction* $\theta \in X$ *if the limit*

$$\lim_{t\to 0} \frac{f(x + t\theta) - f(x)}{t} \tag{1.1.38}$$

exists. If the limit exists for all $\theta \in X$ *we say that* f *is Gâteaux differentiable in* $x \in U$*. Finally, if the limit exists for all* $\theta \in X$ *and for all* $x \in U$ *we simply say that* f *is Gâteaux differentiable.*

We now present the main result of this section, which is also one of the fundamental results in the theory of large deviations. For further details, see [[6], Corollary 4.5.27].

Theorem 1.16 (Gärtner-Ellis)**.** *Let* (μ_t) *be a family of exponentially tight probability measures on a Banach space* $\mathcal{X}$*. Suppose that* $\Lambda(V) = \lim_{t\to\infty} \dfrac{1}{\gamma_t}\Lambda_{\mu_t}(\gamma_t V)$ *is finite valued, Gâteaux differentiable, and lower semicontinuous in* $\mathcal{X}^*$ *with respect to the weak* topology. Then* (μ_t) *satisfies the LDP with the good rate function* Λ^* *and rate* γ_t*.*

To illustrate an application of the Gärtner-Ellis Theorem, we draft a proof of Cramer's Theorem.

Theorem 1.17 (Cramer's Theorem)**.** *Let* $X_1, X_2, \ldots$ *be i.i.d real valued random variables such that* $\varphi(t) = \mathbb{E}[e^{tX_1}] < \infty$ *for all* t*. Further, let* $S_n = \sum_{i=1}^n X_i$*. Then, for every*

$x > \mathbb{E}[X_1]$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nx) = -I(x)\,,$$

where I is the Fenchel-Legendre transformation of $\log\varphi$*, i.e*

$$I(x) = \sup_{t\in\mathbb{R}}[tx - \log\varphi(t)].$$

Proof idea: Let $X_1, X_2, \ldots$ be random variables with moment generating function $\varphi(t) = \mathbb{E}[e^{tX_1}]$ and $S_n = \sum_{i=1}^n X_i$. By Chernoff's bound, we have that $\frac{S_n}{n}$ is exponentially tight. The logarithmic moment generating function is given by

$$\Lambda_{\frac{S_n}{n}}(t) = \log\mathbb{E}\left[e^{t\frac{S_n}{n}}\right] = \log\left(\mathbb{E}\left[e^{\frac{t}{n}X_1}\right]\right)^n = n\log\mathbb{E}\left[e^{\frac{t}{n}X_1}\right] = n\log\varphi\left(\frac{t}{n}\right)$$

and in this way

$$\Lambda(t) = \lim_{n\to\infty}\frac{1}{n}\Lambda_{\frac{S_n}{n}}(tn) = \lim_{n\to\infty}\frac{1}{n}n\varphi\left(\frac{t}{n}n\right) = \log\varphi(t)$$

which is finite for all t by assumption.

The function $\log\varphi(t)$ is Gâteaux differentiable on $\mathbb{R}$ since $\varphi(t)$ is differentiable as a moment generating function, and consequently it is also lower semi-continuous. Since the hypotheses of the Gärtner-Ellis theorem are satisfied, we have that $\frac{S_n}{n}$ satisfies a LDP with a rate function given by

$$I(x) = \sup_{t\in\mathbb{R}}[tx - \log(\varphi(t))]$$

which is the Fenchel-Legendre transform $\Lambda(t)$ and it is exactly the rate function in Cramer's theorem. Therefore

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\frac{S_n}{n}\geq x\right) = -I(x)$$

and we can can conclude. □

Remark 1.18. If (μ_t) is not exponentially tight then the LDP in Theorem 1.16 transforms into a weak LDP.

Remark 1.19. Our goal is to apply the Gärtner-Ellis theorem to the family of random probability measures $(L_t)_{t\geq 0}$ on $\mathbb{Z}^d$. To do so, one would need to determine the dual space of the space of probability measures. However, the space of probability measures is not even a vector space, so that one actually would need to work in an extended space, like the space of all finite measures on $\mathbb{Z}^d$ which has a complicated dual space. In practice this would render the Gärtner-Ellis theorem useless in that context. Fortunately, there is a version of that result that basically states that one can indeed replace the dual space by the space of all continuous bounded functions. For a proof of this result see [17]. Before we state the result we recall that we write $\Lambda(V)$ insted of $\bar{\Lambda}(V)$ whenever the limit in (1.1.34) exists.

Theorem 1.20. *(Random Probability Measures) Let $\mathcal{X}$ be a Polish space and $(\mu_n)_{n\in\mathbb{N}}$ an exponentially tight sequence of $\mathcal{M}_1(\mathcal{X})$-valued random variables. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers with $\gamma_n \to \infty$ as $n \to \infty$. Assume that for every $V \in C_b(\mathcal{X})$, the limit*

$$\Lambda(V) = \lim_{n\to\infty} \frac{1}{\gamma_n} \log \mathbb{E}\left[e^{\gamma_n \langle \mu_n, V\rangle}\right] \tag{1.1.39}$$

exists and is finite.Assume furthermore that the function $\Lambda : C_b(\mathcal{X}) \to \mathbb{R}$ is Gâteaux-differentiable and continuous at zero with respect to pointwise convergence, i.e., for any sequence $(V_n)_{n\in\mathbb{N}}$ in $C_b(\mathcal{X})$ with $\lim_{n\to\infty} V_n(x) = 0$ for all $x \in \mathcal{X}$, we have $\lim_{n\to\infty} \Lambda(V_n) = 0$.

Then the sequence $(\mu_n)_{n\in\mathbb{N}}$ satisfies a Large Deviation Principle with scale γ_n and rate function

$$I(\mu) = \sup_{V\in C_b(\mathcal{X})} \left(\langle \mu, V\rangle - \Lambda(V)\right), \quad \mu \in \mathcal{M}_1(\mathcal{X}).$$

Remark 1.21. As before, without exponential tightness the above result would yield a weak LDP.

1.2 Proof of Theorem 1.10

1.2.1 Preparation

For the proof the Theorem 1.10 we will prove the upper and the lower bound separately.

For the upper bound we note that by Lemma 3.4.1 in [17] one has always for any compact set $K \subseteq \mathcal{M}_1(\mathbb{Z}^d)$ that

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\ell_t \in K) \leq -\inf_{\mu\in K} \bar{\Lambda}^*(\mu)\,, \tag{1.2.1}$$

where for a bounded function V the quantity $\bar{\Lambda}(V)$ was defined in (1.1.37), and $\bar{\Lambda}^*$ denotes its Fenchel-Legendre transform defined in (1.1.35). In this way, we have an upper bound for compact sets with a rate function which apparently differs from the one we desire, but in the sequel we will identify the Fenchel-Legrende transform in (1.2.1) with the function I from Theorem 1.10.

In theory, if we would be able to show that $\bar{\Lambda}(V) = \Lambda(V)$, i.e., that the limit in (1.1.37) exists, then by Theorem 1.20 one should be able to proof the corresponding large deviation lower bound. However, we fail to show the Gâteaux-differentiability of $V \mapsto \Lambda(V)$. The reason for that is that as we will see in Proposition 1.22 the expression $\Lambda(V)$ equals the supremum of the spectrum of the operator $\mathcal{L} + V$. Hence, to show its differentiability in V one needs to resort to perturbation theory of operators. Unfortunately, all existing results in the literature which would imply Gâteaux-differentiability require the supremum to be an isolated point in the spectrum. This however is not even true for the most simplest case in which $\mathcal{L} = \Delta^d$ agrees with the discrete laplacian. Therefore, for the lower bound we adopt a different strategy, which we will explain now. Let $B \subseteq \mathbb{Z}^d$ be a subset of $\mathbb{Z}^d$ containing the origin, we will establish a large deviation principle for the law of $(L_t)_{t\geq 0}$ under the sub-probability measures $\mathbb{P}(\cdot, \mathrm{supp}(L_t) \subseteq B)$. In plain words we only look at trajectories of the random walk that stay inside the set B. The theory of large deviation for such sub-probability measures does not differ from the one for probability measures. In particular we can use Theorem 1.20. We will establish the limit of

$$\Lambda_B(V) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\Big[e^{t\langle \ell_t, V\rangle} \mathbb{1}_{\mathrm{supp}(\ell_t)\subseteq B}\Big] \tag{1.2.2}$$

for any bounded function V with support in B, show that it satisfies all the necessary condition to apply Theorem 1.20, and therefore obtain a LDP of $(L_t)_{t\geq 0}$ under the law $\mathbb{P}(\cdot, \mathrm{supp}(L_t) \subseteq B)$ with some rate function I_B. To finish the argument it then also remains to send B to $\mathbb{Z}^d$.

The first step will be to show that $\Lambda(V)$ and $\Lambda_B(V)$ defined as above exist. This is the content of the following result.

Proposition 1.22. *Let $V : \mathbb{Z}^d \to \mathbb{R}$ be a bounded function. If B is a finite subset of $\mathbb{Z}^d$ containing the origin, then*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{t\langle L_t, V\rangle} \mathbb{1}_{\mathrm{supp}(\ell_t)\subseteq B}] = \sup_{\substack{f\in \ell^2(B),\\ \|f\|_2=1}} \{\langle \mathcal{L}f, f\rangle + \langle Vf, f\rangle\}\,. \tag{1.2.3}$$

Furthermore, we can extend (1.2.3) *to the entire space and obtain*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{t\langle L_t, V\rangle}] = \sup_{\substack{f\in\ell^2(\mathbb{Z}^d),\\ \|f\|_2=1}} \{\langle \mathcal{L}f, f\rangle + \langle Vf, f\rangle\}\,. \tag{1.2.4}$$

Remark 1.23. *Note that the right hand side of* (1.2.4) *coincides with the supremum of the spectrum of $\mathcal{L}+V$, whereas in the finite-dimensional setting the right hand side in* (1.2.3) *is given by the largest eigenvalue of $\mathcal{L}+V$ defined on $\ell^2(B)$.*

Proof. For the proof of this proposition, we start with the following claim.

Claim 1.24. *Let B be a finite subset of $\mathbb{Z}^d$ containing the origin and define*

$$\lambda_1(B,V) := \sup_{\substack{f\in\ell^2(B),\\ \|f\|_2=1}} \{\langle \mathcal{L}f, f\rangle + \langle Vf, f\rangle\}\;.$$

Then $\lambda_1(B,V)$ is the largest isolated eigenvalue of $\mathcal{H}_B = \mathcal{L}+V$ defined on $\ell^2(B)$ in the sense that $\lambda_1(B,V) > \lambda_2(B,V) \geq \lambda_3(B,V) \geq \cdots$. Here, for any $i \in \mathbb{N}$, we denote by $\lambda_i(B,V)$ the i largest eigenvalue of $\mathcal{H}_B$.

To prove the claim, we require the following definition and theorem:

Definition 1.25. *Given a nonnegative $n\times n$ matrix A, we let its rows and columns be indexed in the usual way by $\{1,2,\ldots,n\}$, and we define a directed graph $G(A)$ with vertex set $\{1,2,\ldots,n\}$ by declaring that there is an edge from i to j if and only if $A(i,j)\neq 0$.*

An irreducible matrix is a square nonnegative matrix such that for every i,j there exists $k>0$ such that $A^k(i,j)>0$.

Notice, for any positive integer k, $A^k(i,j)>0$ if and only if there is a path of length k in $G(A)$ from i to j.

Theorem 1.26 (Perron-Frobenius)**.** *Suppose that T is an $n\times n$ non-negative irreducible matrix. Then there exists an eigenvalue r such that:*

(a) r is real and $r>0$;

(b) there exist strictly positive left and right eigenvectors associated with r;

(c) $r\geq|\lambda|$ for any eigenvalue λ with $\lambda\neq r$;

(d) the eigenvectors associated with r are unique up to constant multiples;

(e) If $0\leq B\leq T$ and ρ is an eigenvalue of B, then $|\rho|\leq r$. Moreover, $|\rho|=r$ implies $B=T$;

(f) r is a simple root of the characteristic equation of T.

To use the Perron-Frobenius Theorem and to prove the claim, we first remind the reader that the operator $\mathcal{L}$ has the following matrix representation for all $x, y \in B$

$$\mathcal{L}(x,y) = \begin{cases} a_{x,y}\,, & \text{if } x \neq y\,, \\ -\sum_{z\neq x} a_{x,z}\,, & \text{otherwise.} \end{cases} \tag{1.2.5}$$

In particular we see that $\mathcal{H}_B = \mathcal{L} + V$ might have negative entries on the diagonal which might render the Perron-Frobenius theorem useless for our purposes. However, since B is a finite subset of $\mathbb{Z}^d$ there exists a constant C such that $\mathcal{H}_B + CI_d$ has only non-negative entries, where I_d denotes the identity matrix. Note that the eigenvalues of $\mathcal{H}_B$ and $\mathcal{H}_B + CI_d$ only differ by the additive constant C, and their eigenvectors are preserved. Moreover, by Assumption (A3) in Section 1.1.2 the matrix $\mathcal{H}_B + CI_d$ is irreducible. Thus, the Perron-Frobenius theorem applies, and therefore an eigenvalue r satisfying $r \geq |\lambda|$ for any other eigenvalue with $\lambda \neq r$ exists. If we would know that all eigenvalues were real, then we could conclude. However, the rates $a_{x,y}$ are symmetric by Assumption (A5). Thus, $\mathcal{H}_B + CI_d$ is self-adjoint with respect to the usual inner product in $\ell^2(B)$, and therefore all its eigenvalues are real. This concludes the proof of the claim.

Before we turn to the proofs of (1.2.3) and (1.2.4) we need one additional ingredient. By [16, Section 2.1], for any finite box $B \subseteq \mathbb{Z}^d$ we have the following representation

$$\mathbb{E}_0\left[\exp\left(\int_0^t V(X_s)\mathrm{d}s\right)\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}\delta_z(X_t)\right] = \sum_{k=1}^{|B|}\exp(t\lambda_k(B,V))v_k(B,0)v_k(B,z)\,, \tag{1.2.6}$$

where $v_1(B,\cdot), v_2(B,\cdot), \dots$ form an orthonormal basis of eigenvectors, and $v_k(B,\cdot)$ corresponds to the k largest eigenvalue of $\mathcal{H}_B$. In particular, give $z \in \mathbb{Z}^d$ we denote by $v_k(B,z)$ the value of the eigenvalue $v_k(B,\cdot)$ in z. We are finally in a position to prove Proposition 1.22.

Proof of the lower bound in (1.2.3)*:*

By the Eigenvalue Expansion (1.2.6) we have that for any $z \in B$

$$\begin{aligned}\mathbb{E}_0[e^{t\langle L_t,V\rangle}\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}] &\geq \mathbb{E}_0\left[\exp\left(\int_0^t V(X_s)\mathrm{d}s\right)\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}\delta_z(X_t)\right] \\ &= \exp\left(\lambda_1\left(B,V\right)t\right)\sum_{k=1}^{|B|}\exp\left(\left(\lambda_k(B,V)-\lambda_1(B,V)\right)t\right)v_k(B,0)v_k(B,z)\,.\end{aligned}$$

Applying the log and dividing by t we obtain

$$\frac{1}{t}\log \mathbb{E}_0[e^{t\langle L_t,V\rangle}\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}] \geq \lambda_1\left(B,V\right) + \mathrm{II}\,,$$

where

$$\mathrm{II} = \frac{1}{t}\log\left(\sum_{k=1}^{|B|}\exp((\lambda_k(B,V)-\lambda_1(B,V))t)v_k(B,0)v_k(B,z)\right).$$

Using that $\lambda_1(B,V) > \lambda_k(B,V)$ for all $k>1$, we obtain

$$\mathrm{II} = \log\left(v_1(B,0)v_1(B,z) + \sum_{k=2}^{|B|}\exp((\lambda_k(B,V)-\lambda_1(B,V))t)v_k(B,0)v_k(B,z)\right)^{\frac{1}{t}} \underset{t\to\infty}{\longrightarrow} \log(1).$$

Thus,

$$\liminf_{t\to\infty}\frac{1}{t}\log \mathbb{E}_0[e^{t\langle L_t,V\rangle}\mathbb{1}_{\mathrm{supp}(\ell_t)\subseteq B}] \geq \liminf_{t\to\infty}(\lambda_1(B,V)+\mathrm{II}) \geq \lambda_1(B,V)\,.$$

This finishes the proof of the lower bound.

Proof of the upper bound in (1.2.3)*:*

We have that, summing in the second line below over all possible end points $z\in B$ of the random walk

$$\begin{aligned}
&\limsup_{t\to\infty}\frac{1}{t}\log \mathbb{E}_0\left[e^{t\langle L_t,V\rangle}\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}\right] \\
&= \limsup_{t\to\infty}\frac{1}{t}\log\sum_{z\in B}\mathbb{E}_0\left[\exp\left(\int_0^t V(X_s)\mathrm{d}s\right)\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}\delta_z(X_t)\right] \\
&= \limsup_{t\to\infty}\left[\lambda_1(B,V) + \frac{1}{t}\log\left(\sum_{z\in B}\sum_{k=1}^{|B|}\mathrm{e}^{t(\lambda_k(B,V)-\lambda_1(B,V))}v_k(B,z)v_k(B,0)\right)\right] \\
&\leq \limsup_{t\to\infty}\lambda_1(B,V) + \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{z\in B}\sum_{k=1}^{|B|}\mathrm{e}^{t(\lambda_k(B,V)-\lambda_1(B,V))}v_k(B,z)v_k(B,0)\right).
\end{aligned}$$

Using that the eigenvector form an orthonormal basis of $\ell^2(B)$ and are thus bounded by one, we obtain

$$\begin{aligned}\limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{k=1}^{|B|}\mathrm{e}^{t(\lambda_k(B,V)-\lambda_1(B,V)))}\sum_{z\in B}v_k(B,z)v_k(B,0)\right)\\ \le \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{k=1}^{|B|}\mathrm{e}^{t(\lambda_k(B,V)-\lambda_1(B,V))}|B|\right)\\ \le \limsup_{t\to\infty}\frac{1}{t}\log(|B|^2)=0.\end{aligned}\tag{1.2.7}$$

Consequently,

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}_0\left[\exp\left(\int_0^t V(X_s)\mathrm{d}s\right)\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}\right]\le\lambda_1(B,V).$$

We therefore have proven the upper bound in (1.2.3). Together with the proof of the lower bound we can conclude the proof of Equation (1.2.3).

We will now prove (1.2.4). As above we prove the upper and the lower bound separately.

Proof of the lower bound in (1.2.4)*:* We first note that for any finite subset B of $\mathbb{Z}^d$ we can write

$$\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{E}_0[e^{t\langle L_t,V\rangle}]\ge\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{E}_0[e^{t\langle L_t,V\rangle}\mathbb{1}_{\{\mathrm{supp}(l_t)\subseteq B\}}]=\lambda_1(B,V)\,.\tag{1.2.8}$$

Thus, it remains to investigate what happens in the limit as B tends to $\mathbb{Z}^d$. To that end, recall that

$$\lambda_1(B,V)=\sup_{\substack{g\in L^2(B)\\ \|g\|_2=1}}\langle(\mathcal{L}+V)\,g,g\rangle=\sup_{\substack{g\in L^2(B)\\ \|g\|_2=1}}\left\{\left\langle V,g^2\right\rangle-\frac{1}{2}\sum_{x\sim y}a_{x,y}\Big(g(x)-g(y)\Big)^2\right\}.$$

The second equality follows from a direct calculation using the definition of $\mathcal{L}$. We moreover define,

$$\lambda_1\left(\mathbb{Z}^d,V\right):=\sup_{\substack{g\in L^2(\mathbb{Z}^d)\\ \|g\|_2=1}}\left\{\left\langle V,g^2\right\rangle-\frac{1}{2}\sum_{x\sim y}a_{x,y}\Big(g(x)-g(y)\Big)^2\right\}.\tag{1.2.9}$$

Fixing $\varepsilon>0$, let $f_1\in\ell^2(\mathbb{Z}^d)$ with $\|f_1\|_2=1$ be a near-maximizer, i.e.,

$$\lambda_1(\mathbb{Z}^d,V)-\varepsilon\le\left\langle V,f_1^2\right\rangle-\frac{1}{2}\sum_{x\sim y}a_{x,y}\Big(f_1(x)-f_1(y)\Big)^2.$$

We then define

$$f_{2_B}(x) = \begin{cases} \dfrac{f_1(x)}{\|f_1\|_{l^2(B)}}, \ x \in B \\ 0, \quad x \ \in B^c. \end{cases} \tag{1.2.10}$$

Note that $f_{2_B} \in \ell^2(B)$ and $\|f_{2_B}\|_2 = 1$. When $B \to \mathbb{Z}^d$, we have by the Dominated Convergence Theorem that

$$\lambda_1(\mathbb{Z}^d, V) - \varepsilon \le \lim_{B \to \mathbb{Z}^d} \left\langle V, f_{2_B}^2 \right\rangle - \frac{1}{2} \sum_{\substack{x,y \in \mathbb{Z}^d \\ x \sim y}} a_{x,y} \Big(f_{2_B}(x) - f_{2_B}(y) \Big)^2 \le \lim_{B \to \mathbb{Z}^d} \lambda_1(B, V).$$

Hence, by (1.2.8) we can conclude the proof of the lower bound.

Proof of the upper bound in (1.2.4)*:* Recall the definition of $(u_t)_{t \ge 0}$ in Assumption (A4). We define $\Gamma_t = [-u_t, u_t]^d \cap \mathbb{Z}^d$ and use the notation $X_{[0,t]} = \{X_s : s \in [0,t]\}$ for the trace of the process X between time 0 and time t. We then claim the following:

Claim 1.27. *We have that*

$$\begin{aligned} &\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}_0 \left[\exp\left(\int_0^t V(X_s) \mathrm{d}s \right) \right] \\ &= \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}_0 \left[\exp\left(\int_0^t V(X_s) \mathrm{d}s \right) \mathbb{1}_{\{X_{[0,t]} \subseteq \Gamma_t\}} \right] \end{aligned} \tag{1.2.11}$$

To see that this is true, we write

$$\begin{aligned} &\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}_0 \left[\exp\left(\int_0^t V(X_s) \mathrm{d}s \right) \right] \\ &= \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}_0 \left[\exp\left(\int_0^t V(X_s) \mathrm{d}s \right) \left(\mathbb{1}_{\{X_{[0,t]} \subseteq \Gamma_t\}} + \mathbb{1}_{\{X_{[0,t]} \subsetneq \Gamma_t\}} \right) \right] \\ &= \max\{\mathrm{I}, \mathrm{II}\}, \end{aligned}$$

where

$$\mathrm{I} = \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}_0 \left[\exp\left(\int_0^t V(X_s) \mathrm{d}s \right) \mathbb{1}_{\{X_{[0,t]} \subseteq \Gamma_t\}} \right], \tag{1.2.12}$$

and

$$\mathrm{II} = \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}_0 \left[\exp\left(\int_0^t V(X_s) \mathrm{d}s \right) \mathbb{1}_{\{X_{[0,t]} \subsetneq \Gamma_t\}} \right]. \tag{1.2.13}$$

Note that

$$\begin{aligned}
\mathbb{I}\mathbb{I} &= \limsup_{t\to\infty} \frac{1}{t}\log \mathbb{E}_0\left[\exp\left(\int_0^t V(X_s)\mathrm{d}s\right)\mathbb{1}_{\{X_{[0,t]}\subsetneq \Gamma_t\}}\right] \\
&\leq \limsup_{t\to\infty} \frac{1}{t}\log \mathbb{E}_0\left[\exp(t\|V\|_\infty)\mathbb{1}_{\{X_{[0,t]}\subsetneq \Gamma_t\}}\right] \\
&= \limsup_{t\to\infty}\left(\|V\|_\infty + \frac{1}{t}\log \mathbb{P}_0\left[X_{[0,t]}\subsetneq \Gamma_t\right]\right) = -\infty \quad (\text{by } A_4)\,.
\end{aligned}$$

Hence, the claim follows.

By the above claim and an eigenvalue expansion we see that

$$\begin{aligned}
&\limsup_{t\to\infty} \frac{1}{t}\log \mathbb{E}_0\left[e^{t\langle V, L_t\rangle}\right] \\
&= \limsup_{t\to\infty} \frac{1}{t}\log \sum_{z\in\Gamma_t}\mathbb{E}_0\left[\exp\left(\int_0^t V(X_s)\mathrm{d}s\right)\mathbb{1}_{\{X_{[0,t]}\subseteq \Gamma_t\}}\delta_z(X_t)\right] \\
&= \limsup_{t\to\infty}\left[\lambda_1(\Gamma_t,V) + \frac{1}{t}\log\left(\sum_{z\in\Gamma_t}\sum_{k=1}^{|\Gamma_t|}\mathrm{e}^{t(\lambda_k(\Gamma_t,V)-\lambda_1(\Gamma_t,V))}v_k(\Gamma_t,z)v_k(\Gamma_t,0)\right)\right] \\
&\leq \limsup_{t\to\infty}\lambda_1(\Gamma_t,V) + \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{z\in\Gamma_t}\sum_{k=1}^{|\Gamma_t|}\mathrm{e}^{t(\lambda_k(\Gamma_t,V)-\lambda_1(\Gamma_t,V))}v_k(\Gamma_t,z)v_k(\Gamma_t,0)\right).
\end{aligned}$$

Using that the v_k's all have norm one in $\ell^2(\Gamma_t)$ and are therefore bounded by one, we obtain

$$\begin{aligned}
\limsup_{t\to\infty}\frac{1}{t}\log&\left(\sum_{k=1}^{|\Gamma_t|}\mathrm{e}^{t(\lambda_k(\Gamma_t,V)-\lambda_1(\Gamma_t,V))}\sum_{z\in\Gamma_t}v_k(\Gamma_t,z)v_k(\Gamma_t,0)\right) \\
&\leq \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{k=1}^{|\Gamma_t|}\mathrm{e}^{t(\lambda_k(\Gamma_t,V)-\lambda_1(\Gamma_t,V))}|\Gamma_t|\right) \\
&\leq \limsup_{t\to\infty}\frac{1}{t}\log(|\Gamma_t|^2) = 0.
\end{aligned}$$

Here, we used that $\lambda_1(\Gamma_t,V)\geq\lambda_k(\Gamma_t,V)$ for all $k\in\{1,2,\dots,|\Gamma_t|\}$, and the fact that $|\Gamma_t|\leq|u_t|$ to obtain the last equality. Consequently

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}_0\left[e^{t\langle V,L_t\rangle}\right]\leq\limsup_{t\to\infty}\lambda_1(\Gamma_t,V)\leq\lambda_1(\mathbb{Z}^d,V)\,.$$

The last inequality follows directly from the variational representations of $\lambda_1(\Gamma_t,V)$ and $\lambda_1(\mathbb{Z}^d,V)$. Hence, we can conclude.

□

Remark 1.28. *As already mentioned, by Lemma 3.4,1 in [17], we already have an upper bound for Theorem 1.10, but with a rate function that is the Fenchel-Legendre transform of* $V \mapsto \lambda_1(\mathbb{Z}^d, V)$*. We refer also to* (1.2.1).

As mentioned at the beginning of this section, our goal is to prove a large deviation principle for the empirical measure under the law $\mathbb{P}(\cdot, \text{supp}(L_t) \subseteq B)$. To that end we will now show the differentiability of $V \mapsto \lambda_1(B, V)$.

1.2.2 Differentiability of $V \mapsto \lambda_1(B, V)$

Lemma 1.29. *The application* $V \to \lambda_1(B, V)$ *which maps* $\mathbb{R}^{|B|}$ *into* $\mathbb{R}$ *is differentiable with respect to V. In particular, it is also continuous in zero.*

Proof. Consider the largest eigenvalue of $\mathcal{L} + V$ on $\ell^2(B)$, and note that it is a zero of the characteristic polynomial p of $\mathcal{L} + V$. Since we are interested in differentiability of $\lambda_1(B, V) = \max\{\lambda : p(\lambda, V) = 0\}$ with respect to V it makes sense to define

$$p : \mathbb{R} \times \mathbb{R}^{|B|} \to \mathbb{R}; \quad p(\lambda, V) = \det(\lambda \mathrm{I}_d - \mathcal{L} - V)$$

where I_d is the identity matrix in $\mathbb{R}^{|B|} \times \mathbb{R}^{|B|}$, and study the differentiability, with respect to V, of the map p and make use of the Implicit Function Theorem. In particular, with our notation above, $p(\lambda_1(B, V), V) = 0$. Moreover, by Claim 1.24 we have that $\lambda_1(B, V)$ is an isolated eigenvalue, hence has multiplicity 1. We can therefore write

$$p(\lambda, V) = (\lambda - \lambda_1(B, V))(\lambda - \lambda_2(B, V))^{m_2(V)} \cdots (\lambda - \lambda_\ell(B, V))^{m_l(V)}$$

where $m_i(V)$ is the multiplicity of $\lambda_i(B, V)$ with $2 \leq i \leq l$ and $\sum_{i=2}^{\ell} m_i(V) = |B| - 1$. We note that

$$\frac{\partial}{\partial_\lambda} p(\lambda_1(B, V), V) \neq 0.$$

In fact, we can write

$$p(\lambda, V) = (\lambda - \lambda_1(V)) q(\lambda, V),$$

where q is a polynomial of degree $|B| - 1$ with $q(\lambda_1(V), V) \neq 0$.

Hence,

$$\frac{\partial}{\partial_\lambda} p(\lambda, V) = q(\lambda, V) + (\lambda - \lambda_1(V)) \frac{\partial}{\partial_\lambda} q(\lambda, V).$$

Therefore,

$$\frac{\partial}{\partial_\lambda}p(\lambda_1(B,V),V) = q(\lambda_1(B,V),V) + (\lambda(B,V) - \lambda_1(B,V))\frac{\partial}{\partial_\lambda}q(\lambda_1(B,V),V) \neq 0.$$

By the Implicit Function Theorem there exists an open neighborhood U in $\mathbb{R}^{|B|}$ of V and a unique continuously differentiable function $g : U \to \mathbb{R}$ such that $g(V) = \lambda_1(B,V)$ and $p(g(w),w) = 0$ for all $w \in U$. But we note that g is simply the map that maps $w \in U$ to the largest eigenvalue of $\mathcal{L} + \omega$, and as shown this map is continuously differentiable. Hence, we can conclude. □

Proposition 1.22, and Lemma 1.29 together with Theorem 1.20 yield the following corollary.

Corollary 1.30. *The sequence of empirical measures $(L_t)_{t\geq 0}$ satisfies a Large Deviation principle with rate t and rate function*

$$\widetilde{I}_B(\mu) = \sup_{V\in\mathbb{R}^B} \{\langle V,\mu\rangle - \lambda_1(B,V)\}, \quad \mu \in \mathcal{M}_1(B), \tag{1.2.14}$$

under the sub-probability measures $\mathbb{P}(\cdot, \mathrm{supp}(L_t) \in B)$.

Note that the rate function in (1.2.14) and the Fenchel-Legendre transform of $V \mapsto \lambda_1(\mathbb{Z}^d, V)$ seem to structurally differ from the rate function in Theorem 1.10. In the sequel we will show that this is not the case.

1.2.3 Identification of rate functions

Lemma 1.31. *For all $\mu \in \mathcal{M}_1(\mathbb{Z}^d)$ we have that $\Lambda^*(\mu)$ defined in* (1.1.35) *is equal to $I(\mu)$ defined in* (1.1.14).

Proof. Let $\ell^1(\mathbb{Z}^d)$ and $\ell^\infty(\mathbb{Z}^d)$ be the spaces of the summable and bounded functions mapping from $\mathbb{Z}^d$ into $\mathbb{R}$. It is well known that $\ell^1(\mathbb{Z}^d)^* = \ell^\infty(\mathbb{Z}^d)$. For any $\mu \in \mathcal{M}_1(\mathbb{Z}^d)$ we have that

$$\Lambda^*(\mu) = \sup_{V\in l^\infty(\mathbb{Z}^d)} \left\{\langle V,\mu\rangle - \lambda_1(\mathbb{Z}^d,V)\right\}.$$

As in (1.2.26) we can write

$$\lambda_1(\mathbb{Z}^d,V) = \sup_{\mu\in\mathcal{M}_1(\mathbb{Z}^d)} \{\langle V,\mu\rangle - I(\mu)\} \tag{1.2.15}$$

where we remind the reader that I was defined defined in (1.1.14).

Since $\mathcal{M}_1(\mathbb{Z}^d) \subset L^1(\mathbb{Z}^d)$, we have that

$$L^\infty(\mathbb{Z}^d) = (L^1(\mathbb{Z}^d))^* \subset (\mathcal{M}^1(\mathbb{Z}^d))^*.$$

Then we replace in (1.2.15) the function I by its extension $\hat{I} : \ell^1(\mathbb{Z}^d) \to [0,\infty]$ defined via

$$\hat{I}|_{\ell^1(\mathbb{Z}^d)\setminus\mathcal{M}_1(\mathbb{Z}^d)} \equiv \infty.$$

Then,

$$\lambda_1(\mathbb{Z}^d, V) = \sup_{\mu\in\ell^1(\mathbb{Z}^d)} \left\{\langle V,\mu\rangle - \hat{I}(\mu)\right\} \quad \forall\, V \in \ell^\infty(\mathbb{Z}^d).$$

Hence, $\lambda_1(\mathbb{Z}^d,\cdot) : (l^1(\mathbb{Z}^d))^* \to \mathbb{R}$ is the Fenchel-Legendre transform of $\hat{I} : \ell^1(\mathbb{Z}^d) \to [0,\infty]$. To continue, we will accept the following claim whose proof will be given further below.

Claim 1.32. *The application $\hat{I}$ is lower semicontinuous and convex.*

Hence, by Lemma 3.4.3 in [17] the Fenchel-Legendre transform of $\lambda_1(\mathbb{Z}^d,\cdot)$ is equal to $\hat{I}$, which is defined by

$$\hat{I}(\mu) = \sup_{V\in l^\infty(\mathbb{Z}^d)} \left\{\langle V,\mu\rangle - \lambda_1(\mathbb{Z}^d, V)\right\} \quad \forall \mu \in \ell^1(\mathbb{Z}^d) \tag{1.2.16}$$

and will be proved later.

Since $\hat{I}$ is the extension of I to $\ell^1(Z^d)$ we obtain that for all $\mu \in \mathcal{M}_1(\mathbb{Z}^d)$

$$I(\mu) = \hat{I}(\mu) = \Lambda^*(\mu)\,. \tag{1.2.17}$$

Hence, we can conclude that $I(\mu) = \Lambda^*(\mu)$ for all $\mu \in \mathcal{M}_1(\mathbb{Z}^d)$ as desired. □

Proof of Claim 1.32. Since $\mathcal{M}_1(\mathbb{Z}^d)$ is convex and closed in $\ell^1(\mathbb{Z}^d)$ it is enough to show the lower semicontinuity and convexity of $\mu \mapsto I(\mu)$.

Lower semicontinuity: Note that for any finite subset B of $\mathbb{Z}^d$ the mapping

$$\mu \mapsto \frac{1}{2} \sum_{x,y\in B, x\sim y} a_{x,y}(\sqrt{\mu}(x) - \sqrt{\mu}(y))^2 \tag{1.2.18}$$

is continuous. The claim the follows by the fact that the supremum of lower semi-continuous functions is lower semi-continuous.

Convexity: This follows by a direct computation. Indeed, let $\alpha \in [0,1]$ and $\mu,\nu \in$

$\mathcal{M}_1(\mathbb{Z}^d)$. Then,

$$I(\alpha\mu+(1-\alpha)\nu) = \frac{1}{2}\sum_{x\sim y} a_{x,y}\Big(\sqrt{\alpha\mu(x)+(1-\alpha)\nu(x)} - \sqrt{\alpha\mu(y)+(1-\alpha)\nu(y)}\Big)^2 . \quad (1.2.19)$$

For each $x, y \in \mathbb{Z}^d$ as above we can write the inner term in the sum above as

$$\Big(\alpha\mu(x)+(1-\alpha)\nu(x)\Big)+\Big(\alpha\mu(y)+(1-\alpha)\nu(y)\Big)-2\sqrt{\alpha\mu(x)+(1-\alpha)\nu(x)}\sqrt{\alpha\mu(y)+(1-\alpha)\nu(y)} . \quad (1.2.20)$$

We then have that

$$2\sqrt{\mu(x)\mu(y)\nu(x)\nu(y)} \le \mu(x)\nu(y)+\mu(y)\nu(x) . \quad (1.2.21)$$

Indeed, to see that square both sides above, then it only remains to realize that

$$2\mu(x)\mu(y)\nu(x)\nu(y) \le \mu^2(x)\nu^2(y)+\mu^2(y)\nu^2(x) , \quad (1.2.22)$$

which is true. Hence,

$$\begin{aligned}\sqrt{\alpha\mu(x)+(1-\alpha)\nu(x)}\sqrt{\alpha\mu(y)+(1-\alpha)\nu(y)} &\ge \sqrt{\Big(\alpha\sqrt{\mu(x)\mu(y)}+(1-\alpha)\sqrt{\nu(x)\nu(y)}\Big)^2} \\ &= \alpha\sqrt{\mu(x)\mu(y)}+(1-\alpha)\sqrt{\nu(x)\nu(y)} .\end{aligned} \quad (1.2.23)$$

Plugging the above into (1.2.20) and (1.2.19) is enough to conclude. □

We now come to the case of a finite space. More precisely, we will show the following.

Lemma 1.33. *For all $\mu \in \mathcal{M}_1(B)$ we have that $\widetilde{I}_B(\mu)$, defined in* (1.2.14)*, is equal to $I_B(\mu)$, where for any $\mu \in \mathcal{M}_1(B)$*

$$I_B(\mu) = \frac{1}{2}\sum_{x\sim y} a_{x,y}(\sqrt{\mu}(x)-\sqrt{\mu}(y))^2 . \quad (1.2.24)$$

Proof. To identify $\widetilde{I}_B$ with I_B observe that, after a substitution $\mu = g^2$, the eigenvalue can also be written as follows:

$$\lambda_1(B,V) = \sup_{\mu\in\mathcal{M}_1(B)} \{\langle V,\mu\rangle - I_B(\mu)\} . \quad (1.2.25)$$

Indeed, this follows in the same way as (1.1.14). Observe that the space $\mathbb{R}^B$ is identical to its dual space. When we replace I_B in (1.2.25) by its extension $\hat{I}_B : (\mathbb{R}^B)^* \to [0,\infty]$ defined via $\hat{I}_B|_{(\mathbb{R}^B)^*\setminus\mathcal{M}_1(B)} := \infty$ and extend the supremum to all $\mu \in (\mathbb{R}^B)^*$, (1.2.25) remains valid. This means that

$$\lambda_1(B,V) = \sup_{\mu\in(\mathbb{R}^B)^*} \left\{\langle V,\mu\rangle - \hat{I}_B(\mu)\right\} , \tag{1.2.26}$$

for all $V \in \mathbb{R}^B$. Thus, $\lambda_1(B,\cdot)$ is the Fenchel-Legendre transform of $\hat{I}_B$. Note that $\hat{I}_B$ is lower semi-continuous and convex. The argument is identical to the one of Claim 1.32, and we will not repeat the arguments here. According to the duality principle (see Lemma 3.4.3 in [17]), the Fenchel-Legendre transform of $\lambda_1(B,\cdot)$ is equal to $\hat{I}_B$ on $(\mathbb{R}^B)^*$, i.e., for all $\mu \in (\mathbb{R}^B)^*$,

$$\hat{I}_B(\mu) = \sup_{V\in\mathbb{R}^B} \left\{\langle V,\mu\rangle - \lambda_1(B,V)\right\}. \tag{1.2.27}$$

When we restrict this fact to $\mathcal{M}_1(B)$, we obtain precisely the desired result that $I_B = \hat{I}_B = \widetilde{I}_B$ on $\mathcal{M}_1(B)$. Thus, for all $\mu \in \mathcal{M}_1(B)$ we have that $I_B(\mu) = \widetilde{I}_B(\mu)$ as we intended to show. Hence, we can conclude. □

Our arguments so far show that for all compact subsets K of $\mathcal{M}_1(\mathbb{Z}^d)$ we have the desired large deviation upper bounds. Regarding the lower bound, by Corollary 1.30 and Lemma 1.33 for all $A \subset \mathcal{M}_1(B)$ open, we have that

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(L_t \in A, \mathrm{supp}(l_t) \subseteq B\right) \geq -\inf_{\mu\in A} I_B(\mu).$$

Our goal is to extend this result to the full space, and for this our idea is show that we can approximate any probability measure by a sequence of of measures with support in a finite set with respect to the topology of weak convergence in $\mathcal{M}_1(\mathbb{Z}^d)$. We note to that end that since $\mathbb{Z}^d$ is discrete the weak topology coincides with the topology generated by the total variation norm.

Given $\varepsilon \in (0,1)$, and $\mu \in \mathcal{M}_1(\mathbb{Z}^d)$, let $B \subseteq \mathbb{Z}^d$ be such that $\mu(B^c) \leq \varepsilon$. We define

$$\mu_B(x) = \begin{cases} \dfrac{\mu(x)}{\mu(B)}, & x \in B \\ 0, & \text{otherwise.} \end{cases} \tag{1.2.28}$$

For the following calculations we write

$$\frac{1}{\mu(B)} = \frac{1}{1-\mu(B^c)} = 1 + \frac{\mu(B^c)}{1-\mu(B^c)} .$$

Note that the latter term converges to 1 as $B \to \mathbb{Z}^d$.

For any $A \subseteq \mathbb{Z}^d$,

$$\mu(A) = \mu(A \cap B) + \mu(A \cap B^c) \leq \mu_B(A \cap B) + \mu(B^c) = \mu_B(A) + \mu(B^c)\,, \qquad (1.2.29)$$

where the inequality is a consequence of the definition of μ_B. Moreover

$$\begin{aligned} \mu_B(A) = \frac{\mu(A \cap B)}{1 - \mu(B^c)} &= \mu(A \cap B)\left(1 + \frac{\mu(B^c)}{1 - \mu(B^c)}\right) \\ &\leq \mu(A) + \frac{\mu(B^c)}{1 - \mu(B^c)}. \end{aligned}$$

Consequently, the total variation distance between μ and μ_B is bounded from above by $\frac{\mu(B^c)}{1-\mu(B^c)}$.

We will now complete the proof of Theorem 1.10. More precisely, we will show the following.

Statement: Given $A \subseteq \mathcal{M}_1(\mathbb{Z}^d)$ such that μ is in the interior of A, which we denote by A^{o} , we have

$$\liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(L_t \in A\right) \geq -I(\mu). \qquad (1.2.30)$$

Proof. For all $\nu \in \mathcal{M}_1(B)$ denote by $B_\varepsilon(\nu, B)$ the ball of radius ε in $\mathcal{M}_1(B)$ with center ν. Given A and μ as in the statement, our last calculations show that there is $B \subseteq \mathbb{Z}^d$ and $\varepsilon > 0$ such that $B_\varepsilon(\mu_B, B) \subseteq A^{\mathrm{o}}$. Using the LDP lower bound for $\mathbb{P}\left(\cdot, \mathrm{supp}(l_t) \subseteq B\right)$ we obtain that for all $B \subseteq \mathbb{Z}^d$

$$\begin{aligned} \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t} l_t \in A\right) &\geq \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t} l_t \in B_\varepsilon(\mu, B), \mathrm{supp}(l_t \subseteq B)\right) \\ &\geq -I_B(\mu_B). \end{aligned}$$

It only remains to show that

$$\limsup_{B \uparrow \mathbb{Z}^d} I_B(\mu_B) \leq I(\mu). \qquad (1.2.31)$$

We recall that for $x \in B$:

$$\mu_B(x) = \begin{cases} \mu(x)\left(1 + \dfrac{\mu(B^c)}{1 - \mu(B^c)}\right), x \in B \\ 0, \text{otherwise.} \end{cases}$$

Then, we can write

$$I_B(\mu_B) = \frac{1}{2}\sum_{\substack{x,y\in\mathbb{Z}^d\\ x\sim y}} a_{xy}\left(\sqrt{\mu_B}(x) - \sqrt{\mu_B}(y)\right)^2 = \mathrm{I} + \mathrm{II}$$

where

$$\mathrm{I} = \frac{1}{2}\sum_{\substack{x,y\in B\\ x\sim y}} a_{xy}\left(\sqrt{\mu(x)\left(1+\frac{\mu(B^c)}{1-\mu(B^c)}\right)} - \sqrt{\mu(y)\left(1+\frac{\mu(B^c)}{1-\mu(B^c)}\right)}\right)^2,$$

and

$$\mathrm{II} = \sum_{\substack{x\in B, y\notin B\\ x\sim y}} a_{xy}\left(\sqrt{\mu_B(x)} - \sqrt{\mu_B(y)}\right)^2.$$

We have that

$$\begin{aligned}
\mathrm{I} &= \frac{1}{2}\sum_{\substack{x,y\in B\\ x\sim y}} a_{xy}\left(\sqrt{\mu(x)\left(1+\frac{\mu(B^c)}{1-\mu(B^c)}\right)} - \sqrt{\mu(y)\left(1+\frac{\mu(B^c)}{1-\mu(B^c)}\right)}\right)^2 \\
&= \left(1+\frac{\mu(B^c)}{1-\mu(B^c)}\right)\frac{1}{2}\sum_{\substack{x,y\in B\\ x\sim y}} a_{xy}\left(\sqrt{\mu}(x) - \sqrt{\mu}(y)\right)^2 \\
&\le \left(1+\frac{\mu(B^c)}{1-\mu(B^c)}\right) I(\mu),
\end{aligned}$$

and the last expression converges to $I(\mu)$ as $B \uparrow \mathbb{Z}^d$. For II we have that

$$\begin{aligned}
\frac{1}{2}\sum_{\substack{x\in B, y\notin B\\ x\sim y}} a_{xy}\left(\sqrt{\mu_B(x)} - \sqrt{\mu_B(y)}\right)^2 &= \frac{1}{2}\sum_{\substack{x\in B, y\notin B\\ x\sim y}} a_{xy}\mu_B(x) \\
&= \frac{1}{2}\sum_{x\in B}\mu_B(x)\sum_{y\notin B, y\sim x} a_{xy} \\
&= \frac{1}{2}\sum_{x\in \partial B}\mu_B(x)\sum_{y\notin B, y\sim x} a_{xy} \\
&\le \left(\sup_{x\in\mathbb{Z}^d}\sum_{y\in\mathbb{Z}^d} a_{xy}\right)\sum_{x\in\partial B}\mu_B(x) \\
&= C(a)\mu_B(\partial B) \\
&= C(a)\frac{\mu(\partial B)}{1-\mu(B^c)},
\end{aligned}$$

where ∂B represents the boundary of the set B , and by (A2)

$$C(a) = \sup_{x \in \mathbb{Z}^d} \sum_{y \in \mathbb{Z}^d} a_{xy} < \infty.$$

Since,

$$\lim_{B \to \mathbb{Z}^d} \frac{\mu(\partial B)}{1 - \mu(B^c)} = 0$$

we can conclude. $\square$

In the next Capter we will show A Strong Large Deviation Principle for the Empirical Measure of Random Walks

Chapter 2

A Strong Large Deviation Principle for the Empirical Measure of Random Walks

This chapter is based on an article that was submitted for publication and posted on https://arxiv.org/abs/2409.01290.

2.1 Introduction

Let $X = (X_t)_{t\geq 0}$ be a Markov process on a compact Polish space (Σ, d). In the 1970's, Donsker and Varadhan [8, II] showed that under suitable conditions the empirical measure defined by

$$L_t = \frac{1}{t}\int_0^t \mathbb{1}_{X_s} \mathrm{d}s$$

satisfies a large deviation principle in the space of probability measures equipped with the weak topology. When the state space Σ is *not compact*, the upper bound holds for *compact* sets rather than *closed* sets. Meanwhile, the lower bound remains valid under a certain irreducibility assumption which we will detail further below.

In certain scenarios, this *weak* large deviation principle can be upgraded to a standard (or *strong*) one, thereby recovering the large deviation upper bound for all closed sets. However, such cases do not encompass many natural examples. To address the lack of compactness in applications, a confining drift may be added to the Markov chain (or diffusion) [11], or it may be folded onto a large torus [2, 4, 10]. Folding the Markov process on a large torus is often sufficient when one aims for more crude information like the behaviour of a partition function. Yet, if one is interested in finer properties of the model at hand the above mentioned methods are not enough to account for the lack of a strong large deviation principle. An example one should have in mind here is that many models in statistical mechanics are defined in terms of Gibbs measures. Those often involve a normalization constant (partition function), to normalize the measure. Without a strong large deviation principle one is often only able to analyse the behaviour of the normalization constant, but not of the measure itself.

Recently, Mukherjee and Varadhan [21] introduced a new approach by embedding the space of probability measures on $\mathbb{R}^d$ into a larger space equipped with a certain topology, rendering it a *compact* metric space. Under this new topology, they established a *strong* large deviation principle for the empirical measure of Brownian motion [21, Theorem 4.1], which was then successfully applied to the *Polaron Problem* [3, 15, 21]. The compactification of measures has also proven fruitful in the context of directed polymers, as demonstrated in [1, 5]. For an overview of Large Deviation Theory, readers may consult [6, 7], and for the role of topology in this theory, refer to [23].

2.1.1 Topology of Mukherjee and Varadhan

In this section we will explain the topology of Mukherjee and Varadhan. In the original article [21] the authors constructed a metric space $\widetilde{\mathcal{X}}$ as an enlargement of $\mathcal{M}_1(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$. Since we are interested in a discrete context our construction starts with the space of probability measures $\mathcal{M}_1 = \mathcal{M}_1(\mathbb{Z}^d)$ on $\mathbb{Z}^d$. The changes that need to be done to adapt from the original framework are minor and we outline those changes whenever necessary.

Denote by $\widetilde{\mathcal{M}}_1 = \mathcal{M}_1 \,/\!\sim$ the quotient space of $\mathcal{M}_1$ under the action of $\mathbb{Z}^d$ (as an additive group on $\mathcal{M}_1$). For any $\mu \in \mathcal{M}_1$, its orbit is defined by $\widetilde{\mu} = \{\mu * \delta_x : x \in \mathbb{Z}^d\} \in \widetilde{\mathcal{M}}_1$. For $k \geq 2$, we define $\mathcal{F}_k$ as the space of all functions $f : (\mathbb{Z}^d)^k \to \mathbb{R}$ that are diagonally translation invariant, i.e.,

$$f(u_1 + x, \ldots, u_k + x) = f(u_1, \ldots, u_k) \qquad \forall\, x, u_1, \ldots, u_k \in \mathbb{Z}^d\,,$$

and vanishing once getting away from the diagonal, i.e.,

$$\lim_{\max_{i \neq j} |u_i - u_j| \to \infty} f(u_1, \ldots, u_k) = 0\,.$$

For $k \geq 2$, $f \in \mathcal{F}_k$ and $\mu \in \mathcal{M}_{\leq 1}$ (the space of sub-probability measures on $\mathbb{Z}^d$), we define $\Lambda(f, \mu)$, the integral of f with respect to a product measure of k copies of μ, by

$$\Lambda(f, \mu) := \int f(u_1, \ldots, u_k) \prod_{1 \leq i \leq k} \mu(\mathrm{d}u_i)\,. \tag{2.1.1}$$

Due to the translation invariance of f, the above expression actually only depends on the orbit $\widetilde{\mu}$. The class of test functions that is of special importance is defined via

$$\mathcal{F} = \bigcup_{k \geq 2} \mathcal{F}_k \,.$$

As noted in [21], there exists a countable dense set for each $\mathcal{F}_r$ (under the uniform metric). Taking the union of all of these subsets we obtain a countable set, which we will write as

$$\Big\{ f_r(u_1, \ldots, u_{k_r}) : r \in \mathbb{N} \Big\} \,.$$

The desired compactification of $\widetilde{\mathcal{M}}_{\leq 1}$ (the space of orbits with respect to translations in space of sub-probability measures on $\mathbb{Z}^d$) is the space $\widetilde{\mathcal{X}}$ defined via

$$\widetilde{\mathcal{X}} := \Big\{ \xi = \{\tilde{\alpha}_i\}_{i \in \mathcal{I}} \,:\, \alpha_i \in \mathcal{M}_{\leq 1} \,, \sum_{i \in \mathcal{I}} \alpha_i(\mathbb{Z}^d) \leq 1 \Big\}$$

equipped with the metric

$$\mathbf{D}(\xi_1, \xi_2) := \sum_{r \geq 1} \frac{1}{2^r} \frac{1}{1 + \|f_r\|_\infty} \Big| \sum_{\widetilde{\alpha} \in \xi_1} \Lambda(f_r, \alpha) - \sum_{\widetilde{\alpha} \in \xi_2} \Lambda(f_r, \alpha) \Big| \,.$$

Here, $\mathcal{I}$ denotes an empty, finite or countable index set. Moreover, given $\xi = \{\tilde{\alpha}_i\}_{i \in \mathcal{I}} \in \widetilde{\mathcal{X}}$ we use the notation

$$\Lambda(f, \xi) = \sum_{\widetilde{\alpha} \in \xi} \Lambda(f, \alpha) \,. \tag{2.1.2}$$

Remark 2.1. *We note that in the definition of $\widetilde{\mathcal{X}}$, given any orbit $\tilde{\alpha}_i$ we denote by $\alpha_i \in \mathcal{M}_{\leq 1}$ an arbitrary chosen element of $\tilde{\alpha}_i$. Since all elements have the same total mass, the definition of $\widetilde{\mathcal{X}}$ is well defined.*

Theorem 3.1 in [21] reads as follows.

Theorem 2.2. $\mathbf{D}$ *is a metric on $\widetilde{\mathcal{X}}$.*

The second key result in [21] is its Theorem 3.2 which reads:

Theorem 2.3. *The set of orbits $\widetilde{\mathcal{M}}_1$ is dense in $\widetilde{\mathcal{X}}$. Furthermore, given any sequence $(\widetilde{\mu}_n)_n$ in $\widetilde{\mathcal{M}}_1$, there is a subsequence that converges to a limit in $\widetilde{\mathcal{X}}$. Hence, $\widetilde{\mathcal{X}}$ is a compactification of $\widetilde{\mathcal{M}}_1$. It is also the completion under the metric $\mathbf{D}$ of $\widetilde{\mathcal{M}}_1$.*

Theorem 2.2 follows from the same proof as in [21], Theorem 2.3 however requires some small modifications that we will outline in Section 2.2.1. Before we close this section let us give an illustrative example. Define a measure

$$\mu_n = \frac{1}{2}\mathcal{U}(\{n-1, n, n+1\}) + \frac{1}{3}\mathcal{U}(\{-n-1, -n, -n+1\}) + \frac{1}{6}\mathcal{U}(\{-n, -n+1, \dots, n\}),$$

where $\mathcal{U}(A)$ denotes the uniform distribution on the set A. Then, (μ_n) clearly does not converge in the weak topology. However, one can show that its orbit $(\widetilde{\mu}_n)$ converges in $(\widetilde{\mathcal{X}}, \mathbf{D})$ to

$$\xi = \left\{\frac{1}{2}\mathcal{U}_3(\cdot),\ \frac{1}{3}\mathcal{U}_3(\cdot)\right\}.$$

Here $\mathcal{U}_3(\cdot)$ denotes the orbit of the uniform distribution on a set of three consecutive numbers. We emphasize that ξ indeed consists of two distinct elements. The way to interpret is that the mass of the sequence $(\mu_n)_{n\in\mathbb{N}}$ split into two, and therefore created these two elements of ξ in the limit. We note that the contribution of $\frac{1}{6}\mathcal{U}(\{-n, -n+1, \dots, n\})$ vanishes in the limit as $n \to \infty$.

Remark 2.4. *One might wonder what is gained by introducing this new topology, given that a natural way to compactify the space of probability measures is to use the vague topology. However, this new topology provides information that the vague topology does not. For example, as the above example illustrates, the existence of multiple orbits in the limit implies that the center of mass of the sequence of probability measures was splitting into several parts, with their mutual distances diverging to infinity. Moreover, if the limit does not have total mass one, this simply indicates that part of it has disintegrated. Another important feature is that this topology is naturally applicable to models of statistical mechanics with an inherent shift invariance—see, for instance, [21] for applications to the polaron model.*

2.1.2 Main result

Let $(X_t)_{t\geq 0}$ be a Markov chain on $\mathbb{Z}^d$ with generator $\mathcal{L}$ acting on bounded functions $f : \mathbb{Z}^d \to \mathbb{R}$ via

$$(\mathcal{L}f)(x) = \sum_{y\in\mathbb{Z}^d} a_{x,y}[f(y) - f(x)]\,. \tag{2.1.3}$$

We denote the domain of $\mathcal{L}$ by $\mathcal{D}$ and we assume that the rates $(a_{x,y})_{x,y\in\mathbb{Z}^d}$ satisfy the following six assumptions.

(B1) The process $X = (X_t)_{t\geq 0}$ is a Feller process.

(B2) $\sup_{x\in\mathbb{Z}^d} \sum_{y\in\mathbb{Z}^d} a_{x,y} < \infty$.

(B3) For all $x, y \in \mathbb{R}^d$ there exist $n \in \mathbb{N}$ and a sequence of points $x = x_0, x_1, \dots, x_n = y$ such that for all $i \in \{0, 1, \dots, n-1\}$ we have that $a_{x_i, x_{i+1}} > 0$.

(B4) For all $c, x, y \in \mathbb{Z}^d$ we have that $a_{x,y} = a_{x+c,y+c}$.

(B5) There exists a positive function $(u_t)_{t \geq 0}$ satisfying $\lim_{t\to\infty} \frac{1}{t} \log u_t = 0$ such that

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\Big(\sup_{0 \leq s \leq t} |X_s| \geq u_t \Big) = -\infty.$$

(B6) The rates are symmetric in the sense that for all $x, y \in \mathbb{Z}^d$ one has that $a_{x,y} = a_{y,x}$.

The third condition of course simply states that $(X_t)_{t\geq 0}$ is an irreducible Markov chain.

Remark 2.5. *We remark that the above conditions are in particular satisfied for the symmetric simple random walk and for any symmetric random walk with finite range if it is irreducible and if the rate to jump from x to y depends only on the distance of x and y.*

Remark 2.6. *The attentive reader might have noticed that the above conditions almost coincide with those in Chapter 1, the only difference is* (B4) *above. This condition is very natural in the present context, since shift invariance is a definining feature of the topology that we are considering. On the more technical side, assumption* (B4) *above will guarantee that the rate function that we will obtain in our main result Theorem 2.7 below is well defined in the sense that it does not depend on the chosen representant of the orbit.*

Next, we define the quantity of interest, namely, the *normalized occupation measure* $(L_t)_{t\geq 0}$ defined via

$$L_t(A) = \frac{1}{t} \int_0^t \mathbb{1}_A(X(s))\, \mathrm{d}s\,, \qquad A \subseteq \mathbb{Z}^d\,. \tag{2.1.4}$$

Note that defined in this way, for each $t \geq 0$, the quantity L_t is a random probability measure on $\mathbb{Z}^d$. We recall that in [9], under the assumptions (B1)–(B3), it was shown that $(L_t)_{t\geq 0}$ satisfies a weak large deviation principle with respect to the usual weak convergence of probability measures with rate function I defined on $\mathcal{M}_1(\mathbb{Z}^d)$ via

$$I(\mu) = -\inf_{\substack{u>0\\ u\in\mathcal{D}}} \int \Big(\frac{\mathcal{L}u}{u}\Big)(x)\mu(\mathrm{d}x) = \Big\|(-\mathcal{L})^{\frac{1}{2}}\sqrt{\mu}\Big\|^2 = \sum_{x,y\in\mathbb{Z}^d} a_{x,y}(\sqrt{\mu}(x) - \sqrt{\mu}(y))^2, \tag{2.1.5}$$

where the norm appearing above is on the space $\ell^2(\mathbb{Z}^d)$. The second equality is thanks to Assumption (B6) and is a consequence of [8, I, Theorem 5], while the last equality follows from a straightforward computation. We note at this point that the above definition also makes sense for sub-probability measures $\mu \in \mathcal{M}_{\leq 1}(\mathbb{Z}^d)$ and we will make use of that without further mentioning it. In the forthcoming Subsection 2.2.2 we will compare our assumptions to the one in [8].

To proceed we extend the rate function I to $\widetilde{\mathcal{X}}$. We define $\widetilde{I} : \widetilde{\mathcal{X}} \to \mathbb{R}$ for $\xi \in \widetilde{\mathcal{X}}$ via

$$\widetilde{I}(\xi) = \sum_{\widetilde{\alpha} \in \xi} I(\alpha)\,, \tag{2.1.6}$$

where α is an arbitrary representative of $\widetilde{\alpha}$. This makes sense because we will show in Lemma 2.13 that, as a consequence of Assumption (B4), the rate function I depends on μ only through its orbit. Finally denote by $\widetilde{L}_t$ (the orbit of the probability measure L_t) the embedding of L_t into $\widetilde{\mathcal{X}}$ and denote its law by Q_t.

Our main result then reads as follows.

Theorem 2.7. *Assume that Assumptions* (B1)–(B6) *hold. Then, the family of measures* $(Q_t)_{t\geq 0}$ *defined on* $\widetilde{\mathcal{X}}$ *satisfy a large deviation principle with rate function* $\widetilde{I}$ *and rate* t*.*

Before we come to the end of that section we will shortly discuss our assumptions. As mentioned above, Assumptions (B1)–(B3) will guarantee that the family of empirical measures satisfies a weak large deviation principle in the usual weak topology in $\mathcal{M}_1$. Note that the topology of Mukherjee-Varadhan is applied to contexts in which the model at hand exhibits a shift invariance in space. In particular, the large deviation principle in $\widetilde{\mathcal{X}}$ can then be seen as an extension of the weak large deviation principle on $\mathcal{M}_1$. It is therefore natural to consider models in which the rate function I governing the weak large deviation principle actually only depends on the orbit of the measure at hand. Our Assumption (B4) guarantees exactly that. Regarding Assumption (B5) it should be satisfied in almost all examples and simply states that it is very costly for the random walk to travel exponentially large distances. This will come in handy for a coarse graining argument that will be used to establish the large deviation upper bound. Finally, we believe that Theorem 2.7 should hold also without the Assumption (B6). However, without Assumption (B6) the second equality in (2.1.5) does not hold which makes the analysis way more delicate.

2.1.3 Structure of the paper

In Section 2.2 we present the proofs of Theorems 2.3 and 2.7. To that end, we first show in Section 2.2.2 that our assumptions yield a weak large deviation principle in the usual weak topology in $\mathcal{M}_1$. In Section 2.2.1 we prove Theorem 2.3. Then, in Section 2.2.2 we prove that our assumptions imply a weak large deviation principle in the usual weak topology in $\mathcal{M}_1(\mathbb{Z}^d)$. In the remaining subsection of Section 2.2 we complete the proof of Theorem 2.7. Finally, in Section 2.3 we provide two applications of our main result.

2.2 Proofs

In this section we prove Theorems 2.7. Recall that to establish a large deviation principle three properties need to be established. The lower semi-continuity of the rate function, the lower bound for open sets and the upper bound for closed sets. Our first task is to check that our hypothesis indeed imply the hypothesis of Donsker and Varadhan [9], which, in turn, implies a weak LDP, that is, a large deviations principle for compact sets in the usual topology of $\mathcal{M}_1(\mathbb{Z}^d)$.

2.2.1 Adaptation of the proof of Theorem 2.3 and auxiliary results

We start with the proof of Theorem 2.3, which is an adaptation of the proof of [21, Theorem 3.2]. It is divided in two steps:

Step 1. $\widetilde{\mathcal{M}}_1$ is dense in $\widetilde{\mathcal{X}}$.

Step 2. Any sequence $(\mu_n)_n$ has a subsequence that converges to some $\xi \in \widetilde{\mathcal{X}}$.

We start with Step 1. Let $\xi = \{\widetilde{\alpha}_i : i \in \mathcal{I}\} \in \widetilde{\mathcal{X}}$. We are going to show that there exists a sequence of $(\widetilde{\mu}_n)$ of orbits in $\widetilde{\mathcal{M}}_1(\mathbb{Z}^d)$ such that $\widetilde{\mu}_n$ converges to ξ in $\widetilde{\mathcal{X}}$. As argued in [21, Theorem 3.2] we can assume that ξ is a finite collection of orbits of sub-probability measures. To obtain a better understanding of how one should define the sequence of orbits in $\widetilde{\mathcal{M}}_1$, recall that after the example following Theorem 2.3 we already commented that ξ having several elements should be a consequence of the fact that the approximating sequence of orbits has a center of mass which splits into several parts. Therefore, when constructing the desired sequence $\tilde{\mu}_n$, one needs to create these centers of masses that drift apart. For any $i \in \mathcal{I}$ define $p_i = \alpha_i(\mathbb{Z}^d)$ and choose spatial points $(a_i)_{i\in\mathcal{I}} = (a_i(n))_{i\in\mathcal{I}}$ such that $\inf_{i\neq j} |a_i(n) - a_j(n)| \to \infty$ as $n \to \infty$. Also, let ν_n denote the uniform distribution of $[-n, n]^d \cap \mathbb{Z}^d$, i.e.,

$$\nu_n(x) = \frac{1}{(2n+1)^d} \begin{cases} 1, & \text{if } x \in [-n,n]^d \cap \mathbb{Z}^d\,, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2.1}$$

We then note that for any $f \in \mathcal{F}_k$,

$$\lim_{n\to\infty} \int f(x_1,\ldots,x_k) \prod_{i=1}^{k} \nu_n(\mathrm{d}x_i) = 0\,.$$

Indeed, this is a consequence of the fact that (ν_n) totally disintegrates in the sense that, for any $r > 0$,

$$\lim_{n\to\infty} \sup_{x\in\mathbb{Z}^d} \nu_n(B(x,r)) = 0\,,$$

and Lemma 2.3 in [21] whose proof carries over *mutatis mutandis* to the present context. We then define

$$\mu_n = \sum_{i\in\mathcal{I}} \alpha_i * \delta_{a_i} + \Big(1 - \sum_{i\in\mathcal{I}} p_i\Big)\nu_n\,.$$

Since $f \in \mathcal{F}_k$ vanishes whenever the distance of two coordinates tends to infinity and $\inf_{i\neq j} |a_i - a_j| \to \infty$ we see that

$$\lim_{n\to\infty} \int f(x_1,\ldots,x_k) \prod_{i=1}^{k} \mu_n(\mathrm{d}x_i) = \sum_{i\in\mathcal{I}} \int f(x_1,\ldots,x_k) \prod_{j=1}^{k} \alpha_i(\mathrm{d}x_j)\,.$$

This shows that the sequence of orbits $(\widetilde{\mu}_n)$ converges to ξ in $\widetilde{\mathcal{X}}$, since $\Lambda(f,\mu_n) \to \sum_{i\in\mathcal{I}} \Lambda(f,\alpha_i)$ for every $f \in \mathcal{F}_k$. Hence, $\widetilde{\mathcal{M}}_1$ is dense in $\widetilde{\mathcal{X}}$.

The second step of the proof in [21] shows that every sequence $(\widetilde{\mu}_n)$ has a converging subsequence in $\widetilde{\mathcal{X}}$. The proof in the discrete context works without any major adaptations, relying fundamentally on the following lemma ([21, Lemma 2.2]).

Lemma 2.8. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of sub-probability measures that converges vaguely to a sub-probability measure α, then each μ_n can be written as $\mu_n = \alpha_n + \beta_n$. Here $(\alpha_n)_{n\in\mathbb{N}}$ converges in the weak topology to α and $(\beta_n)_{n\in\mathbb{N}}$ converges in the vague topology to zero. Moreover, α_n and β_n can be chosen to have disjoint supports and such that the support of each α_n is compact.*

The above formulation of Lemma 2.8 is a bit more general than in its original version in [21] but its statement follows directly from the proof of [21, Lemma 2.2]. We also refer to [12] where this was already noted. Before we formulate the next result we will provide one more definition.

Definition 2.9. *Two sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ of sub-probability measures on $\mathbb{Z}^d$ are widely separated if for any function $W : \mathbb{Z}^d \to \mathbb{R}$ that vanishes at infinity*

$$\lim_{n\to\infty} \int W(x-y)\,\alpha_n(\mathrm{d}x)\beta_n(\mathrm{d}y) = 0\,.$$

We note that, apart from the discrete scenario, the above definition differs from the one given in [21, Subsection 2.4], which requires W to be strictly positive. However, as one can see in [21, Lemma 2.4], this condition of strict positivity can be dropped, and our definition is equivalent to theirs. Lemma 2.8 together with the proof of the second step in [21, Theorem 3.2] then yields the following corollary.

Corollary 2.10. *Let $(\widetilde{\mu}_n)$ be a sequence in $\widetilde{\mathcal{X}}$, converging to $\xi = \{\widetilde{\alpha}_i\}_{i\in I} \in \widetilde{\mathcal{X}}$. Then there exists a sub-sequence which we also denote by $(\widetilde{\mu}_n)$ such that for any $k \le |\mathcal{I}|$ we can write*

$$\mu_n = \sum_{i=1}^{k} \alpha_{n,i} + \beta_n$$

so that

- *$(\alpha_{n,i})_{n\in\mathbb{N}}, i = 1, \dots, k$, and $(\beta_n)_{n\in\mathbb{N}}$ are sequences of sub-probability measures in $\mathbb{Z}^d$;*
- *for each $i = 1, \dots, k$ there are sequences $(a_{n,i})_{n\in\mathbb{N}}$ of elements of $\mathbb{Z}^d$ such that*

$$\alpha_{n,i} * \delta_{a_{n,i}} \Rightarrow \alpha_i \in \widetilde{\alpha}_i, \qquad \text{as } n \to \infty\,,$$
$$\lim_{n\to\infty} \min_{i\neq j} |a_{n,i} - a_{n,j}| = \infty\,,$$

  and $(\beta_n)_{n\in\mathbb{N}}$ is widely separated from each $(\alpha_{n,i})_{n\in\mathbb{N}}$;

- *The supports of $\alpha_{n,1}, \alpha_{n,2}, \dots, \beta_n$ are all disjoint and for each i there exists a sequence $(R_{n,i})_{n\in\mathbb{N}}$ tending to infinity such that*

$$\mathrm{supp}(\alpha_{n,i}) \subset B(-a_{n,i}, R_{n,i})$$

  and

$$\mathrm{supp}(\beta_n) \subset \Big[\cup_i B(-a_{n,i}, R_{n,i})\Big]^{\complement}.$$

Here $B(-a_{n,i}, R)$ denotes the ball in the Euclidean norm with radius R and center $-a_{n,i}$.

We finish this section with one more result which is a consequence of the proof in [21, Theorem 3.2].

Lemma 2.11. *Assume that the sequence $(\widetilde{\alpha}_n)$ converges in $\widetilde{\mathcal{M}}_1$ to $\widetilde{\alpha} \in \widetilde{\mathcal{M}}_1$. Then $(\widetilde{\alpha}_n)$ converges in $\widetilde{\mathcal{X}}$ to $\widetilde{\alpha}$.*

The proof of this result is a consequence of the analysis of the case $q = p$ (those indices are defined therein) in [21, Proof of Theorem 3.2] and it will be omitted here.

2.2.2 Weak LDP

We start by recalling the assumptions in [9], keeping the original notation. Denote by $p(t,x,\mathrm{d}y)$ the transition probability of the Markov process $(X_t)_{t\geq 0}$ under consideration whose state space is a Polish space X. We will denote its law by $\mathbb{P}_x$ when started at x, and the corresponding expectation will be denoted by $\mathbb{E}_x$. The following four assumptions were put in place:

(DV1) The transition probability $p(t,x,\mathrm{d}y)$ is Feller, having a density $p(t,x,y)$ with respect to a reference measure $\beta(\mathrm{d}y)$.

(DV2) Denote by $(T_t)_{t\geq 0}$ the semi-group of X. Define

$$B_0 = \{f \in C(X) : \lim_{t\to 0} \|T_t f - f\| = 0\},$$

and

$$B_{00} = \{f \in C(X) : \lim_{t\to 0} \sup_x \mathbb{E}_x\Big[|f(X_t) - f(x)|\Big] = 0\}.$$

Moreover, for $\alpha \in \mathcal{M}_1(X)$ such that $I(\alpha) < \infty$, $k \in \mathbb{N}$ and fixed $f_1, \dots, f_k \in B_{00}$ define a neighbourhood $\mathcal{N}_\alpha$ of α in $\mathcal{M}_1(X)$ via

$$\mathcal{N}_\alpha = \left\{\mu \in \mathcal{M}_1(X) : \left|\int_X f_j(x)[\mu(dx) - \alpha(dx)]\right| < \varepsilon, 1 \leq j \leq k\right\}.$$

It was then assumed that $p(t,x,\mathrm{d}y)$ is such that every neighboorhod of $\alpha \in \mathcal{M}_1(X)$ contains a neighborhood of the form $\mathcal{N}_\alpha$.

(DV3) For any $\mu \in \mathcal{M}_1(X)$ and any bounded measurable function f, there exists a sequence $(f_n)_{n\in\mathbb{N}} \in B_0^{\mathbb{N}}$ such that $\|f_n\| \leq \|f\|$ for all $n \in \mathbb{N}$ and such that $f_n \to f$ almost everywhere with respect to μ.

(DV4) For every $x \in X, \sigma > 0$ and $E \subset X$ such that $\beta(E) > 0$, it holds that

$$\int_0^\infty \exp(-\sigma t) p(t,x,E)\mathrm{d}t > 0\,,$$

where $p(t,x,E)$ is defined via

$$p(t,x,E) = \mathbb{P}_x(X_t \in E)\,.$$

Below we argue why our assumptions (B1), (B2) and (BA) indeed imply (DV1), (DV2), (DV3) and (DV4):

(B1) **implies** (DV1) The Feller property is clear by assumption and the reference measure β is in our case simply the counting measure since our state space is discrete. Hence, (DV1) holds.

(B2) **implies** (DV2) We note that the usual topology of weak convergence of probability measures is induced by testing against continuous, bounded functions. Therefore, it remains to argue that B_{00} coincides with $C_b(\mathbb{Z}^d)$ in our case. This is sufficient, since the weak convergence of probability measures is generated by testing against continuous bounded functions. Fix $f \in C_b(\mathbb{Z}^d)$, assume that $X_0 = x$ and define the first jump time via

$$T_1 = \inf\{t > 0 : X_t \neq x\}\,.$$

Since T_1 has an exponential distribution with parameter $\sum_y a_{xy}$ we can estimate

$$\begin{aligned}\mathbb{E}_x\Big[|f(X_t) - f(x)|\Big] &= \mathbb{E}_x\Big[|f(X_t) - f(x)|\mathbb{1}_{\{T_1 > t\}}\Big] + \mathbb{E}_x\Big[|f(X_t) - f(x)|\mathbb{1}_{\{T_1 \leq t\}}\Big] \\ &\leq 0 + 2\|f\|_\infty \mathbb{P}_x(T_1 \leq t) \\ &= 2\|f\|_\infty \sum_y a_{xy} \int_0^t \exp\Big\{-s\sum_y a_{xy}\Big\} ds\,.\end{aligned}$$

By (B2) the latter goes to zero uniformly in x as $t \to 0$. Thus, $f \in B_{00}$ and (DV2) holds.

(B2) **implies** (DV3) Since the state space $\mathbb{Z}^d$ is a discrete space, we have that every function $f : \mathbb{Z}^d \to \mathbb{R}$ is continuous, and if this function is bounded then by the above arguments it clearly belongs to B_0. Thus, it is sufficient to take $f_n = f$, and (DV3) follows.

(B3) **implies** (DV4) This is a direct consequence of the fact that (B3) states that $(X_t)_{t\geq 0}$ is irreducible.

2.2.3 Lower bound

In this section we give the proof of the large deviation lower bound. More precisely, we will show the following result.

Proposition 2.12 (Lower bound)**.** *For any open set* $G \in \widetilde{\mathcal{X}}$

$$\liminf_{t\to\infty} \frac{1}{t} \log Q_t(G) \geq -\inf_{\xi\in G} \widetilde{I}(\xi). \tag{2.2.2}$$

For the proof we will need some properties of the rate function that are provided in the next lemma.

Lemma 2.13. *The rate function I defined in* (2.1.5) *on the space $\mathcal{M}_{\leq 1}(\mathbb{Z}^d)$ is translation invariant, i.e., for any $c \in \mathbb{Z}^d$ one has that $I(\mu) = I(\mu * \delta_c)$, it is homogeneous of degree 1, i.e., for any $\lambda \geq 0$ one has that $I(\lambda\mu) = \lambda I(\mu)$, and it is convex. Consequently it is also sub-additive.*

Lemma 2.14. *Let $\xi \in \widetilde{\mathcal{X}}$ with $\widetilde{I}(\xi) < \infty$. Then, there exists a sequence (ξ_n) in $\widetilde{\mathcal{M}}_1(\mathbb{Z}^d)$ such that*

$$\limsup_{n\to\infty} \widetilde{I}(\xi_n) \leq \widetilde{I}(\xi). \tag{2.2.3}$$

The proofs of the above lemmas are deferred to the end of that section. We now provide the proof of Proposition 2.12.

Proof of Proposition 2.12. Note that to show (2.2.2) it is enough to prove that, given $\xi \in \widetilde{\mathcal{X}}$ with $\widetilde{I}(\xi) < \infty$,

$$\liminf_{t\to\infty} \frac{1}{t} \log Q_t(U) \geq -\widetilde{I}(\xi) \tag{2.2.4}$$

for any open neighbourhood U of ξ. First of all note that for any open neighbourhood U of ξ the set $U \cap \widetilde{\mathcal{M}}_1$ is open in $\widetilde{\mathcal{M}}_1$. Indeed, given any $\theta \in U \cap \widetilde{\mathcal{M}}_1$ and any sequence $(\xi_n)_{n\in\mathbb{N}}$ in $\widetilde{\mathcal{M}}_1$ that converges to θ with respect to the topology in $\widetilde{\mathcal{M}}_1$ it follows from Lemma 2.11 that ξ_n converges also with respect to the topology of $\widetilde{\mathcal{X}}$ to θ. Thus, since U is open in $\widetilde{\mathcal{X}}$ eventually $\xi_n \in U$, and since it is a sequence in $\widetilde{\mathcal{M}}_1$ we have therefore that $\xi_n \in U \cap \widetilde{\mathcal{M}}_1$ provided that n is sufficiently large. Thus, $U \cap \widetilde{\mathcal{M}}_1$ is open in $\widetilde{\mathcal{M}}_1$ as claimed. Since $\widetilde{L}_t \in \widetilde{\mathcal{M}}_1$ we have that

$$\mathbb{P}(\widetilde{L}_t \in U) = \mathbb{P}(\widetilde{L}_t \in U \cap \widetilde{\mathcal{M}}_1).$$

Define

$$\begin{aligned} q : \mathcal{M}_1 &\to \widetilde{\mathcal{M}}_1 \\ u &\mapsto q(u) = \widetilde{u}\,, \end{aligned}$$

as the canonical map from $\mathcal{M}_1$ to $\widetilde{\mathcal{M}}_1$. Note that q is a continuous mapping, so that $q^{-1}(U \cap \widetilde{\mathcal{M}}_1)$ is an open set. Now consider a sequence $(\xi_n)_{n\in\mathbb{N}}$ as in Lemma 2.14. Then for all n sufficiently large we have that $\xi_n \in U \cap \widetilde{\mathcal{M}}_1$. For those n we can estimate

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\widetilde{L}_t \in U) &= \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}(L_t \in q^{-1}(U \cap \widetilde{\mathcal{M}}_1)) \\ &- \inf\{I(\mu) \,:\, \mu \in q^{-1}(U \cap \widetilde{\mathcal{M}}_1) \\ &\geq -I(\xi_n) \\ &= -\widetilde{I}(\xi_n).\end{aligned}$$

Here, we used in the inequality the fact that $(L_t)_{t\geq 0}$ satisfies the large deviation lower bound in the usual weak topology of probability measures together with the fact that by Lemma 2.13 the rate function I only depends on the orbit of the measure. The last equality uses that ξ_n as an element of $\widetilde{\mathcal{X}}$ contains only one element. To conclude it suffices to send n to infinity. □

We now come to the proof of Lemma 2.13.

Proof of Lemma 2.13. We first show that I is translation invariant. We write, using Assumption (B4) in the last equality

$$\begin{aligned}\sum_{x,y\in\mathbb{Z}^d} a_{x,y}(\sqrt{\mu}(x+c) - \sqrt{\mu}(y+c))^2 &= \sum_{x,y\in\mathbb{Z}^d} a_{x-c,y-c}(\sqrt{\mu}(x) - \sqrt{\mu}(y))^2 \\ &= \sum_{x,y\in\mathbb{Z}^d} a_{x,y}(\sqrt{\mu}(x) - \sqrt{\mu}(y))^2\,.\end{aligned}$$

Homogeneity, and convexity are direct consequences of the definition of I. The sub-additivity is then a immediate consequence of the homogeneity and the convexity. Hence, we can conclude. □

We now come to the proof of Lemma 2.14.

Proof of Lemma 2.14. By our analysis at the beginning of Section 2.2.1 the sequence $\widetilde{\mu}_n$ approximating ξ in $\widetilde{\mathcal{X}}$ can be chosen to be of the form

$$\mu_n = \sum_{i\in J_n} \alpha_i * \delta_{a_i} + \left(1 - \sum_{i\in J_n} p_i\right)\nu_n\,,$$

where ν_n is defined in (2.2.1). Here, given $\xi = \{\alpha_i \,:\, i \in \mathcal{I}\}$ as at the beginning of the proof we denote by (J_n) a suitable chosen sequence of subsets of $\mathcal{I}$ that tend to I in the limit as n tends to infinity (note that in Section 2.2.1 the first step we detailed was for the case in which $\mathcal{I}$ was finite). By Lemma 2.13 the rate function I is sub-additive on

$\mathcal{M}_{\leq 1}(\mathbb{Z}^d)$, thus

$$\begin{aligned} I(\mu_n) &= I\Bigg(\sum_{i\in J_n}(\alpha_i * \delta_{a_i})) + \Big(1-\sum_{i\in J_n}p_i\Big)\nu_n\Bigg) \\ &\leq \sum_{i\in J_n} I(\alpha_i * \delta_{a_i}) + \Big(1-\sum_{i\in J_n}p_i\Big)I(\nu_n) \\ &= \sum_{i\in J_n} I(\alpha_i) + \Big(1-\sum_{i\in J_n}p_i\Big)I(\nu_n)\,. \end{aligned}$$

Using that ν_n is the uniform measure on $[-n,n]^d\cap\mathbb{Z}^d$ and therefore puts vanishing mass on the boundary of $[-n,n]^d\cap\mathbb{Z}^d$ and Assumption (B2) show that $\lim_{n\to\infty} I(\nu_n)=0$. Since, by the definition of $\widetilde{I}$ in (2.1.6)

$$\sum_{i\in J_n} I(\alpha_i) \leq \widetilde{I}(\xi)\,,$$

we can conclude. □

2.2.4 Upper bound

Proposition 2.15 (Upper bound)**.** *For any closed set F in $\widetilde{\mathcal{X}}$, we have that*

$$\limsup_{t\to\infty}\frac{1}{t}\log Q_t(F) \leq -\inf_{\xi\in F}\widetilde{I}(\xi). \tag{2.2.5}$$

Before diving into the proof we present some definitions and observations that will be useful for our purposes. Let $\mathcal{U}$ be the space of functions of the form $u=c+v$ where v is a smooth non-negative function with compact support on $\mathbb{R}^d$ and $c>0$ is a constant. Let φ be a function satisfying $0\leq\varphi(x)\leq 1$, $\varphi(x)=1$ inside the unit ball and $\varphi(x)=0$ outside the ball of radius 2. For any $k\geq 1, R>0, u_1,...,u_k\in\mathcal{U}$ and $a_1,...,a_k\in\mathbb{Z}^d$ and $c>0$ we consider the function

$$g(x)=g(k,R,c,a_1,...,a_k,x)=c+\sum_{i=1}^k u_i(x+a_i)\varphi\Big(\frac{x+a_i}{R}\Big) \tag{2.2.6}$$

and define $F:\mathcal{M}_1\to\mathbb{R}$ by setting

$$F(u_1,\dots,u_k,c,R,\mu)=\sup_{\substack{a_1,\dots,a_k:\\ \inf_{i\neq j}|a_i-a_j|\geq 4R}}\int_{\mathbb{Z}^d}\frac{-\mathcal{L}(g(x))}{g(x)}\mu(\mathrm{d}x). \tag{2.2.7}$$

Since the last expression depends only on the image $\widetilde{\mu}$ of μ, we can extend the above definition to $\widetilde{\mathcal{M}}_1$ via

$$\widetilde{F}(u_1, u_2, \ldots, u_k, c, R, \widetilde{\mu}) := F(u_1, u_2, \ldots, u_k, c, R, \mu).$$

We will prove first that $\widetilde{F}(\cdot)$ grows only sub-exponentially as $t \to \infty$.

Lemma 2.16. *For any $k \geq 1, R > 0,\ u_1, \ldots, u_k \in \mathcal{U}$ and $c > 0$,*

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}\{\exp\{t\widetilde{F}(u_1, \ldots, u_k, c, R, \widetilde{L}_t)\}\} \leq 0. \tag{2.2.8}$$

Proof. The proof proceeds in two steps. In the first step we show that the result follows if there were no supremum over $a_1, a_2, \ldots, a_k$ in the definition of $\widetilde{F}$. To be more precise we will actually show that

$$\mathbb{E}\bigg\{\exp\bigg\{\int_0^t \frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\bigg\}\bigg\}$$

is bounded from above uniformly in the choice of $a_1, \ldots, a_k \in \mathbb{Z}^d$. By the Feynman-Kac formula, the function

$$\Psi(t, x) = \mathbb{E}_x\bigg\{g(X_t)\exp\bigg\{\int_0^t \frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\bigg\}\bigg\}$$

is a solution to the initial value problem

$$\begin{cases} \dfrac{\partial}{\partial t}\Psi(t, x) = (\mathcal{L}\Psi)(t, x) - \dfrac{(\mathcal{L}g)(x)}{g(x)}\Psi(t, x)\,, \\ \Psi(0, x) = g(x). \end{cases}$$

Moreover, it is known that the above problem has a unique solution, see for example [14] for a detailed discussion about it. Note that $\Psi(t, x) = g(x)$ solves the above heat equation. Furthermore, by definition (2.2.6) we have that $g(x) \geq c$. Therefore, by the uniqueness of the solution

$$g(x) = \mathbb{E}_x\bigg\{g(X_t)\exp\bigg\{\int_0^t \frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\bigg\}\bigg\} \geq c\mathbb{E}_x\bigg\{\exp\bigg\{\int_0^t \frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\bigg\}\bigg\},$$

and consequently

$$\mathbb{E}_x\bigg\{\exp\bigg\{\int_0^t \frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\bigg\}\bigg\} \leq \frac{g(x)}{c}. \tag{2.2.9}$$

Since g is uniformly bounded from above in $a_1, \ldots, a_k \in \mathbb{Z}^d$ we obtain (2.2.8) in the case in which one removes the supremum in the definition of $\tilde{F}$. We come to the second step of the proof, namely we will deal with the supremum over $(a_1, ..., a_k)$ inside the expectation. To that end we will use a coarse graining argument. First note that if $\sup_{s\in[0,t]}|X_s| = r_t$ once any $|a_i|$ exceeds $r_t + 2R$ it will no longer affect the value of g, since in this case we have that $|\frac{X_s+a_i}{R}| \geq 2$ and φ is supported in the ball of radius 2. Thus, we can limit each a_i to the ball of radius $r_t + 2R$. By Assumption (B5) there exists a positive function $(u_t)_{t\geq 0}$ with $\lim_{t\to\infty}\frac{1}{t}\log u_t = 0$ and a function $M(t)$ tending to infinity as $t \to \infty$ such that

$$\mathbb{P}\Big(\sup_{0<s\leq t}|X_s| \geq u_t\Big) \leq \exp(-tM(t)),$$

for all $t \geq 0$. We will see that this will allow us to ignore the contributions of those trajectories satisfying $\sup_{0<s\leq t}|X_s| \geq u_t$. We can assume that $2R \leq u_t$, so that we can restrict the a_i's to balls of radius $2u_t$. Moreover, the function

$$\frac{(-\mathcal{L}g)(x)}{g(x)}$$

is bounded from above by some constant c_2. This is a consequence of Assumption (B2) and the fact that g is a bounded function which additionally is bounded away from zero by c. With the above considerations in mind we can now estimate

$$\begin{aligned}
&\mathbb{E}\left\{\exp\left\{\sup_{\substack{a_1,\ldots,a_k\in\mathbb{Z}^d\\ |a_i-a_j|\geq 4R\ \forall i\neq j}}\int_0^t\frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\right\}\right\}\\
&\leq \mathbb{E}\left\{\exp\left\{\sup_{\substack{|a_1|\leq 2u_t,\ldots,|a_k|\leq 2u_t\\ |a_i-a_j|\geq 4R\ \forall i\neq j}}\int_0^t\frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\right\}\right\} + \exp(c_2t)\mathbb{P}\Big(\sup_{0\leq s\leq t}|X_s|\geq u_t\Big)\\
&\leq \mathbb{E}\left\{\sum_{\substack{|a_1|\leq 2u_t,\ldots,|a_k|\leq 2u_t,\\ |a_i-a_j|\geq 4R}}\exp\left\{\int_0^t\frac{(-\mathcal{L}g)(X_s)}{g(X_s)}\mathrm{d}s\right\}\right\} + \exp(c_2t - tM(t))\\
&\leq (2u_t)^{dk}\sup_{\substack{(a_1,\ldots,a_k)\in\mathbb{Z}^d\\ |a_i-a_j|\geq 4R}}\mathbb{E}\left\{\exp\left\{\int_0^t\frac{-\mathcal{L}(g(X_s))}{g(X_s)}\mathrm{d}s\right\}\right\} + \exp(c_2t - tM(t)).
\end{aligned}$$

Note that by the first step, the expectation above can be bounded by $g(k, R, c, a_1, \ldots, a_k, x)/c$ which actually is bounded uniformly in the choice of $a_1, \ldots, a_k$. Hence, taking the logarithm above, dividing by t, sending t to infinity shows that the left hand side in (2.2.8) is at most zero.

□

Lemma 2.17. *Let $(\widetilde{\mu}_n)$ be a sequence in $\widetilde{\mathcal{X}}$ which converges to $\xi = \{\widetilde{\alpha}_j\} \in \widetilde{\mathcal{X}}$. For any*

$k \in \mathbb{N}, i = 1, \dots, k$ *and* $u_{i,R}(x) = u_i(x)\varphi\left(\frac{x}{R}\right)$*, where* $u_i \in \mathcal{U}$*, we have*

$$\begin{aligned}\liminf_{n\to\infty} \widetilde{F}(u_1, \dots, u_k, c, R, \widetilde{\mu}_n) \geq \sup_{\alpha_1,\dots,\alpha_k\in\xi} \sum_{i=1}^{k} \sup_{b\in\mathbb{Z}^d} \int \frac{(-\mathcal{L}u_{i,R})(x)}{c + u_{i,R}(x)} \alpha_i(dx + b) \\ =: \widetilde{\Lambda}(\xi, R, c, u_1, \dots, u_k). \qquad (2.2.10)\end{aligned}$$

Before diving into the proof of the above lemma we need to formulate a technical result, which will be useful in the proof of Lemma 2.17 and for the forthcoming results.

Lemma 2.18. *Let* u *be a non-negative function with compact support and* $(\alpha_n)_{n\in\mathbb{N}}$ *be a sequence of sub-probability measures such that*

$$\lim_{n\to\infty} \operatorname{dist}(\operatorname{supp}(\alpha_n), \operatorname{supp}(u)) = \infty\,, \qquad (2.2.11)$$

where $\operatorname{dist}(A, B)$ *denotes the distance in the* ℓ^1*-norm between two sets* $A, B \subseteq \mathbb{Z}^d$*. Under the above condition the following holds for all* $c > 0$*,*

$$\lim_{n\to\infty} \int \frac{-(\mathcal{L}u)(x)}{c + u(x)} \alpha_n(\mathrm{d}x) = 0\,. \qquad (2.2.12)$$

We first prove Lemma 2.17 accepting the validity of Lemma 2.18. The proof of the latter will be given afterwards.

Proof. Fix $b \in \mathbb{Z}^d$ and $\alpha_1, \alpha_2, \dots, \alpha_k \in \xi$. We will follow the second step in the proof of [21, Theorem 3.2]. By Corollary 2.10 we can find $\ell \geq k$ and a subsequence μ_{n_k} that we will suppress from the notation such that we have the decomposition

$$\mu_{n_k} = \sum_{j=1}^{\ell} \alpha_{n,j} + \beta_n$$

so that for a suitable choice of $a_{n_{i,j}}$ that satisfy

$$\lim_{n\to\infty} |a_{n,i} - a_{n,j}| = \infty, \quad \text{for } i \neq j,$$

one additionally has that

$$\alpha_{n,j} * \delta_{a_{n,j}+b} \to \alpha_j * \delta_b\,, \qquad \text{for} \qquad j \in \{1, 2, \dots, k\}\,, \qquad (2.2.13)$$

and such that the supports of the $\alpha_{n,j}$'s and β_n's are all disjoint.

For n large enough we have that $|a_{n,i} - a_{n,j}| \geq 4R$ so that the supports of the $\{u_{i,R}(\cdot - a_{n,i})\}_i$ are mutually disjoint. Now define

$$\text{supp}(u_{i,R}(\cdot - a_{n,i})) = U_i, \ \forall i = 1, \ldots, k$$

and fix $x \in \mathbb{Z}^d$. Since integrating a function against $\alpha_{n,j} * \delta_{a_{n,j}+b}$ is the same as integrating $f(\cdot - a_{n,j} - b)$ against $\alpha_{n,j}$, we have with $a_i = -a_{n,i}$ the two following cases:

- There exist $j \in \{1, \ldots, k\}$ such that $x \in U_j$ and consequently $x \notin U_i \ \ \forall i \neq j$. In this case we have that

$$\begin{aligned}\frac{-\mathcal{L}g(k,R,c,a_1,\ldots a_k,x)}{g(k,R,c,a_1,\ldots,a_k,x)} &= \frac{\sum_{i=1}^k -\mathcal{L}u_{i,R}(x-a_{n,i})}{c+\sum_{i=1}^k u_{i,R}(x-a_{n,i})}\\ &= \frac{\sum_{i=1}^k -\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})}\\ &= \sum_{i=1}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})}.\end{aligned}$$

- $\forall j \in \{1, \ldots, k\}$ we have that $x \notin U_j$. Then, $u_{j,R}(x - a_i) = 0$ for all j and in this case we have that

$$\begin{aligned}\frac{-\mathcal{L}g(k,R,c,a_1,\ldots a_k,x)}{g(k,R,c,a_1,\ldots,a_k,x)} &= \frac{\sum_{i=1}^k -\mathcal{L}u_{i,R}(x-a_{n,i})}{c+\sum_{i=1}^k u_{i,R}(x-a_{n,i})}\\ &= \frac{\sum_{i=1}^k -\mathcal{L}u_{i,R}(x-a_{n,i})}{c}\\ &= \sum_{i=1}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})}.\end{aligned}$$

Summing up the two contributions plus some elementary calculations yield that

$$\begin{aligned}\frac{-\mathcal{L}g(k,R,c,a_1,\ldots a_k,x)}{g(k,R,c,a_1,\ldots,a_k,x)} &= \sum_{j=1}^k\sum_{i=1}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})}\mathbb{1}_{\{x\in U_j\}} + \sum_{i=1}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})}\mathbb{1}_{\{x\notin \cup_{j=1}^k U_j\}}\\ &= \sum_{j=1}^k \mathbb{1}_{\{x\in U_j\}}\left\{\sum_{i=1,i\neq j}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})} + \frac{-\mathcal{L}u_{j,R}(x-a_{n,j})}{c+u_{j,R}(x-a_{n,j})}\right\}\\ &+ \mathbb{1}_{\{x\notin \cup_{j=1}^k U_j\}}\sum_{i=1}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})}\\ &= \sum_{j=1}^k \mathbb{1}_{\{x\in U_j\}}\sum_{i=1,i\neq j}^k \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})} + \sum_{j=1}^k \frac{-\mathcal{L}u_{j,R}(x-a_{n,j})}{c+u_{j,R}(x-a_{n,j})}\\ &+ \sum_{j=1}^k\sum_{i=1,i\neq j}^k \mathbb{1}_{\{x\in U_j\}}\frac{\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})}.\end{aligned}$$

To obtain the last inequality we added and substracted

$$\sum_{j=1}^{k}\Big\{\sum_{i=1,i\neq j}^{k}\mathbb{1}_{\{x\in U_i\}}\frac{\mathcal{L}u_j(x-a_{n,j})}{c+u_j(x-a_{n,j})}+\mathbb{1}_{\{x\notin\bigcup_{i=1}^{k}U_i\}}\frac{-\mathcal{L}u_j(x-a_{n,j})}{c+u_j(x-a_{n,j})}\Big\}. \tag{2.2.14}$$

For $i\neq j$, we have that

$$\begin{aligned}\int_{U_j}\left|\frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})}\right|\mu_n(dx)&=\int_{U_j}\left|\frac{\sum_{y\in U_i}a_{x,y}[u_{i,R}(x-a_{n,i})-u_{i,R}(y-a_{n,i})]}{c+u_{j,R}(x-a_{n,j})}\right|\mu_n(dx)\\&=\int_{U_j}\frac{\sum_{y\in U_i}a_{x,y}u_{i,R}(y-a_{n,i})}{c+u_{j,R}(x-a_{n,j})}\mu_n(dx). \qquad (2.2.15)\end{aligned}$$

Using that u is non-negative and bounded from above and that each μ_n is a sub-probability measure we can estimate the above by some proportionality constant times

$$\sum_{y:\,|y-x|\geq d(U_i,U_j)}a_{x,y}\,. \tag{2.2.16}$$

The latter goes to zero as $n\to\infty$ by (B2). By the same argument, we have that

$$\int_{U_j}\sum_{\substack{i=1\\ i\neq j}}^{k}\frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})}\mu_n(dx)$$

tends to zero as $n\to\infty$. Thus,

$$\begin{aligned}\liminf_{n\to\infty}\int\frac{-\mathcal{L}g(x)}{g(x)}\mu_n(dx)=\liminf_{n\to\infty}\Bigg(&\int\sum_{j=1}^{k}\mathbb{1}_{\{x\in U_j\}}\sum_{i=1,i\neq j}^{k}\frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})}\mu_n(dx)\\&+\int\sum_{j=1}^{k}\frac{-\mathcal{L}u_{j,R}(x-a_{n,j})}{c+u_{j,R}(x-a_{n,j})}\mu_n(dx) \qquad (2.2.17)\\&+\int\sum_{j=1}^{k}\sum_{\substack{i=1\\ i\neq j}}^{k}\mathbb{1}_{\{x\in U_j\}}\frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})}\,\mu_n(dx)\Bigg).\end{aligned}$$

This can be lower bounded by

$$\liminf_{n\to\infty} \int \sum_{j=1}^{k} \mathbb{1}_{\{x\in U_j\}} \sum_{i=1, i\neq j}^{k} \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{j,R}(x-a_{n,j})} \mu_n(dx) \tag{2.2.18}$$

$$+ \liminf_{n\to\infty} \int \sum_{j=1}^{k} \frac{(-\mathcal{L}u_{j,R})(x-a_{n,j})}{c+u_{j,R}(x-a_{n,j})} \mu_n(dx) \tag{2.2.19}$$

$$+ \sum_{j=1}^{k} \liminf_{n\to\infty} \int \sum_{\substack{i=1\\ i\neq j}}^{k} \mathbb{1}_{\{x\in U_j\}} \frac{-\mathcal{L}u_{i,R}(x-a_{n,i})}{c+u_{i,R}(x-a_{n,i})} \mu_n(dx) \tag{2.2.20}$$

$$= \sum_{j=1}^{k} \int \frac{-\mathcal{L}u_{j,R}(x)}{c+u_{j,R}(x)} \alpha_j(dx)\,. \tag{2.2.21}$$

Here, we used the considerations after Equation (2.2.15) to show that (2.2.18) and (2.2.20) converge to zero, and we used Lemma 2.18 to deal with (2.2.19). Since this argument works for any collection $\alpha_1, \dots, \alpha_k \in \xi$ and since we can redefine $a'_{n,i} := a_{n,i} + b_i$, using (2.2.13) we can conclude that

$$\liminf_{n\to\infty} \widetilde{F}(u_1, \dots, u_k, c, R, \widetilde{\mu}_n) \geq \sup_{\alpha_1,\dots,\alpha_k\in\xi} \sum_{j=1}^{k} \sup_{b\in\mathbb{R}^d} \int \frac{-\mathcal{L}u_{j,R}(x)}{c+u_{j,R}(x)} \alpha_j(dx+b)$$

and the lemma is proved. □

Proof of Lemma 2.18. Let $u \in \mathcal{U}$ and let (α_n) be a sequence of sub-probability measures. We have that

$$(\mathcal{L}u)(x) = \sum_y a_{x,y}[u(y) - u(x)].$$

By (2.2.11) we have that $\text{supp}(\alpha_n) \cap \text{supp}(u) = \varnothing$, for n large enough. For such n we can write

$$\int \frac{-(\mathcal{L}u)(x)}{c+u(x)} \alpha_n(dx) = \int \frac{-(\mathcal{L}u)(x)}{c} \alpha_n(dx) \leq \frac{1}{c} \sup_{x\in\mathbb{Z}^d} \sum_{y\in\text{supp}(\alpha_n)} a_{x,y}\,.$$

The latter term goes to zero by Assumption (B2) and the hypothesis of the lemma.

Lemma 2.19. *For* $\widetilde{\Lambda}$ *defined in* (2.2.10) *and* $\widetilde{I}$ *defined in* (2.1.6), *we can write*

$$\widetilde{I}(\xi) = \sup_{\substack{R,c>0,\ k\in\mathbb{N}\\ u_1,\dots,u_k\in U}} \widetilde{\Lambda}(\xi, R, c, u_1, \dots, u_k)$$

Proof. By the classical rate function I defined in (2.1.5) for any $\alpha \in \mathcal{M}_{\leq 1}(\mathbb{R}^d)$ we can identify I as

$$I(\alpha) = \sup_{\substack{c>0\\ u\in\mathcal{U}}} \int \frac{-\mathcal{L}u(x)}{c+u(x)}\alpha(\mathrm{d}x)\,,$$

see also [8, I]. Therefore, for every $k \in \mathbb{N}$

$$\sup_{\substack{R,c>0\\ u_1,\ldots,u_k\in\mathcal{U}}} \widetilde{\Lambda}(\xi, R, c, u_1, \ldots, u_k) = \sup_{\alpha_1,\ldots,\alpha_k\in\xi} \sum_{i=1}^{k} I(\alpha_i)\,.$$

Since

$$\widetilde{I}(\xi) = \sup_{k\in\mathbb{N}} \sup_{\alpha_1,\ldots,\alpha_k\in\xi} \sum_{i=1}^{k} I(\alpha_i)$$

we can conclude. $\square$

We now come to the proof of Proposition 2.15.

Proof. Since $\widetilde{\mathcal{X}}$ is a compact set, to prove (2.2.5) it is enough to prove that if $\xi \in \widetilde{\mathcal{X}}$ and $B_\delta(\xi)$ is a ball of radius δ around ξ then

$$\limsup_{\delta\to 0} \limsup_{t\to\infty} \frac{1}{t} \log Q_t(B_\delta(\xi)) \leq -\widetilde{I}(\xi).$$

In fact, let $C \subseteq \widetilde{\mathcal{X}}$ be a closed set. Given $\varepsilon > 0$ and $\xi \in \widetilde{\mathcal{X}}$ there is $\delta(\xi) = \delta(\xi, \varepsilon)$ such that

$$\limsup_{t\to\infty} \frac{1}{t} \log Q_t(B_{\delta(\xi)}(\xi)) \leq -\widetilde{I}(\xi) + \varepsilon\,. \tag{2.2.22}$$

Writing

$$C \subset \bigcup_{\xi\in C} B_{\delta(\xi)}(\xi)\,,$$

since C as a closed set in the compact space $\widetilde{\mathcal{X}}$ is compact as well there exist $\xi_1, \xi_2, \ldots, \xi_n \in C$ such that

$$C \subset \bigcup_{i=1}^{n} B_{\delta(\xi_i)}(\xi_i).$$

Therefore

$$\begin{aligned}\limsup_{t\to\infty}\frac{1}{t}\log Q_t(C) &\le \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{i=1}^{n} Q_t(B_{\delta(\xi_i)}(\xi_i))\right)\\ &= \max_{i=1}^{n}\left\{\limsup_{t\to\infty}\frac{1}{t}\log Q_t(B_{\delta(\xi_i)}(\xi_i))\right\}\\ &\le \max_{i=1}^{n}\{-\widetilde{I}(\xi_i)\}+\varepsilon\\ &\le -\inf_{\xi\in F}\widetilde{I}(\xi)+\varepsilon.\end{aligned}$$

Here, we used the fact that for any two positive sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ one has that

$$\limsup_{n\to\infty}\frac{1}{n}\log(a_n+b_n)=\max\left\{\limsup_{n\to\infty}\frac{1}{n}\log a_n, \limsup_{n\to\infty}\frac{1}{n}\log b_n\right\} \tag{2.2.23}$$

to obtain the first equality above. To obtain the second inequality we made use of (2.2.22). Fix $k\in\mathbb{N}$, $u_1,\dots,u_k\in\mathcal{U}$, $c,R>0$ and write $F(\widetilde{\mu})=F(u_1,\dots,u_k,c,R,\widetilde{\mu})$, where $F(u_1,\dots,u_k,c,R,\cdot):\widetilde{\mathcal{M}}_1(\mathbb{R}^d)\to\mathbb{R}$ is defined in (2.2.7). Then

$$\begin{aligned}Q_t(B_\delta(\xi))=\mathbb{P}(\widetilde{L}_t\in B_\delta(\xi)) &\le \mathbb{P}(F(\widetilde{L}_t)\in F(B_\delta(\xi)))\\ &\le \mathbb{P}(F(\widetilde{L}_t)\ge \inf_{\widetilde{\mu}\in B_\delta(\xi)} F(\widetilde{\mu}))\\ &\le \exp\Big(-t\inf_{\widetilde{\mu}\in B_\delta(\xi)}F(\widetilde{\mu})\Big)\mathbb{E}\Big[\exp(tF(\widetilde{L}_t))\Big].\end{aligned}$$

From Lemma 2.16 we have that

$$\begin{aligned}\limsup_{t\to\infty}\frac{1}{t}\log Q_t(B_\delta(\xi)) &\le -\inf_{\widetilde{\mu}\in B_\delta(\xi)}F(\widetilde{\mu})+\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}\Big[\exp(tF(\widetilde{L}_t))\Big]\\ &\le -\inf_{\widetilde{\mu}\in B_\delta(\xi)}F(\widetilde{\mu})\,.\end{aligned}$$

Then, by Lemma 2.17 for all $k\in\mathbb{N}$, $u_1,\dots,u_k\in\mathcal{U}$, $c,R>0$

$$\limsup_{\delta\to 0}\limsup_{t\to\infty}\frac{1}{t}\log Q_t(B_\delta(\xi))\le -\Lambda(u_1,\dots,u_k,C,R,\xi)\,.$$

Recall the definition of $\widetilde{I}$ in (2.1.6) Using Lemma 2.19 in the last equality below, we obtain

$$\begin{aligned}\limsup_{\delta\to 0}\limsup_{t\to\infty}\frac{1}{t}\log Q_t(B_\delta(\xi)) &\le \inf_{\substack{R,c>0,\ k\in\mathbb{N}\\ u_1,\ldots,u_k}}(-\Lambda(u_,\ldots,u_k,c,R,\xi))\\ &= -\sup_{\substack{R,c>0,\ k\in\mathbb{N}\\ u_1,\ldots,u_k}}\Lambda(u_,\ldots,u_k,c,R,\xi)\\ &= -\widetilde{I}(\xi)\,.\end{aligned}$$

This proves the result. □

2.2.5 Lower semicontinuity

Lemma 2.20. *If $\xi_n \to \xi$ in $\widetilde{\mathcal{X}}$, then*

$$\liminf_{n\to\infty}\widetilde{I}(\xi_n) \ge \widetilde{I}(\xi).$$

Proof. The proof of this lemma follows largely along the lines of the proof of Lemma 2.17. In particular, we will use the identification

$$\widetilde{I}(\xi) = \sup_{\substack{R,c>0,\ k\in\mathbb{N}\\ u_1,\ldots,u_k\in U}}\widetilde{\Lambda}(\xi,R,c,u_1,\ldots,u_k)$$

established in Lemma 2.19. Therefore, we fix $R,c>0$ and $u_1,\ldots,u_k\in\mathcal{U}$ and we define $u_{i,R}$ as in the formulation of Lemma 2.17.

Let $(\widetilde{\mu}_n)$ be a sequence in $\widetilde{\mathcal{X}}$, converging to $\xi = \{\widetilde{\alpha}_i\}_{i\in I} \in \widetilde{\mathcal{X}}$. We suppose that $\liminf_{n\to\infty}\widetilde{I}(\mu_n) \le l$ for some $0\le l<\infty$ and we will prove that $\widetilde{I}(\xi)\le l$.

Initially consider the case where for each n one has that $\widetilde{\mu}_n$ consists of a single orbit, so that $\widetilde{I}(\mu_n) = I(\mu_n)$. Restricting to a subsequence if necessary, for some $k\ge 1$ we can write by Corollary 2.10

$$\mu_n = \sum_{i=1}^{k}\alpha_{n,i} + \beta_n$$

so that

- $\alpha_{n,i},\ i=1,\ldots,\ k$, and β_n are sequences of sub-probability measures in $\mathbb{Z}^d$;
- for each $i=1,\ldots,k$ there are sequences $a_{n,i},\ i=1,\ldots,k$, in $\mathbb{Z}^d$ that can be chosen

such that

$$\alpha_{n,i} * \delta_{a_{n,i}} \Rightarrow \alpha_i \in \widetilde{\alpha}_i,\ n \to \infty$$
$$\lim_{n\to\infty} \min_{i\neq 1} |a_{n,i} - a_{n,j}| = \infty$$

and (β_n) is widely separated from each $(\alpha_{n,i})$.

- The supports of $\alpha_{n,1}, \alpha_{n,2}, \dots, \beta_n$ are all disjoint and for each i there exists a sequence $(R_{n,i})_n$ tending to infinity such that

$$\text{supp}(\alpha_{n,i}) \subset B(-a_{n,i}, R_{n,i})$$

and

$$\text{supp}(\beta_{n,i}) \subset \Big[\cup_i B(-a_{n,i}, R_{n,i})\Big]^{\complement}.$$

Next, we note that by the compactness of the support of $u_{i,R}$ the support of each $u_{i,R}(\cdot - a_{n,i})$ is contained in $B(a_{n,i}, R)$. We see that we are in the same setting as in the proof of Lemma 2.17. Therefore, as in that proof we can conclude that

$$l \geq \liminf_{n\to\infty} I(\mu_n) \geq \liminf_{n\to\infty} \sum_{i=1}^{k} \int \frac{-\mathcal{L}u_{i,R}(\cdot - a_{n,i})}{c + u_{i,R}(\cdot - a_{n,i})} \mathrm{d}\mu_n = \sum_{i=1}^{k} \int \frac{-\mathcal{L}u_{i,R}}{c + u_{i,R}} \mathrm{d}\alpha_i \,.$$

Note that the right hand side almost equals the right hand side in the first line of (2.2.10). The crucial difference is that there are no suprema over $b \in \mathbb{Z}^d$, and $R, c > 0$. However, as in the proof of Lemma 2.17 we can obtain the shift by b by simply shifting each of the $a_{n,i}$'s. Taking on both sides the supremum over all $R, c > 0$, $k \in \mathbb{N}$ and $u_1, \dots, u_k$ allows with the help of Lemma 2.19 to conclude that

$$\ell \geq \sum_{i=1}^{k} I(\alpha_i)\,. \tag{2.2.24}$$

Hence, this shows the result in the case where each $\widetilde{\mu}_n$ consists of a single orbit. Finally, to treat the general case, that is when $\widetilde{\mu}_n$ has possibly more than one orbit, the idea is the same as in the last paragraph of [21, proof of Lemma 4.2]. □

2.3 Applications of Theorem 2.7

In this section we present two applications of Theorem 2.7.

2.3.1 Application 1

We consider the same assumptions as in Section 2.1.2 and we define the function

$$V(x,y) = \mathbb{1}_{\{x=y\}} \tag{2.3.1}$$

on $\mathbb{Z}^d \times \mathbb{Z}^d$. For any element $\xi \in \widetilde{\mathcal{X}}$ we then consider

$$\Lambda(V,\xi) = \sum_{\alpha\in\xi} \int V(x,y)\, \alpha(dx)\alpha(dy)\,, \tag{2.3.2}$$

which is a continuous functional in ξ with respect to the metric $\mathbf{D}$ defined on $\widetilde{\mathcal{X}}$. Note that

$$\Lambda(V,L_t) = \frac{1}{t^2}\int_0^t ds \int_0^t du\, \mathbb{1}_{\{X_s=X_u\}}\,, \tag{2.3.3}$$

which is simply the (rescaled by t^2) intersection local time of the random walk X. To continue define a new rate function I' on $\mathbb{R}$ via

$$I'(y) = \inf\{\widetilde{I}(\xi) : \xi \in \widetilde{\mathcal{X}}, \Lambda(V,\xi) = y\}\,. \tag{2.3.4}$$

Intersection local times are an object of intensive studies in the literature, see for instance [19], and with the techniques developed here, we are able to provide a strong large deviation principle as a direct corollary of Theorem 2.7 together with the contraction principle.

Theorem 2.21. *The process of rescaled intersection local times $(\Lambda(V,L_t))_{t\geq 0}$ satisfies a strong large deviation principle with rate function I' and rate t.*

2.3.2 Application 2

Our next application is related to a discrete version of the polaron problem. Consider a bounded strictly positive function $V \in \mathcal{F}_2$ taking its maximum in $(0,0)$ such that $V(0,0) > \sum_y a_{0,y}$, and define

$$Z_t = \mathbb{E}_0\Big[\exp\Big(t\int V(x,y)L_t(dx)L_t(dy)\Big)\Big] = \mathbb{E}_0\Big[\exp\Big(\frac{1}{t}\int V(X_s,X_u)\,\mathrm{d}s\,\mathrm{d}u\Big)\Big]\,. \tag{2.3.5}$$

Remark 2.22. The definition of Z_t is somewhat related to the Polaron problem [11] which corresponds to the choice $d=3$, $V(x,y)=\frac{1}{|x-y|}$ and where the underlying process is a Brownian motion instead of a random walk. We see that this function does not fall into the class of functions considered here. The reason is simply that this choice of function renders Z_t meaningless. One could possibly circumvent this by studying an appropriate choice of approximations of the potential of the Polaron problem. The drawback would be that this would translate to sequences of empirical measures that depend on an additional approximation parameter N. In this work we prefer to refrain from it. Finally, let us comment the technical assumption $V(0,0) > \sum_y a_{0,y}$. This assumption will guarantee, as we will see below, that there is disappearance of mass. To be more precise, it will guarantee that in Theorem 2.23 the supremum on the right hand side is over the space of probability measures and not over the space of sub-probability measures.

It follows from Laplace-Varadhan's Lemma [6, Theorem 4.3.1] and Theorem 2.7 that

$$\lim_{t\to\infty} \frac{1}{t} \log Z_t = \sup_{\xi\in\widetilde{\mathcal{X}}} \left\{ \Lambda(V,\xi) - \widetilde{I}(\xi) \right\} =: \lambda\,, \tag{2.3.6}$$

where $\Lambda(V,\xi)$ was introduced in (2.1.2). We will analyze the above variational formula in the sequel and will show that there is a maximizer and that this maximizer is an orbit of a probability measure. We will start with the existence of the maximizer. Consider a sequence $(\xi_n)_n$ in $\widetilde{\mathcal{X}}$ such that

$$\lim_{n\to\infty} \left\{ \Lambda(V,\xi_n) - \widetilde{I}(\xi_n) \right\} = \lambda\,.$$

By the compactness of $\widetilde{\mathcal{X}}$ we can extract a subsequence of $(\xi_n)_n$ that converges to an element $\xi \in \widetilde{\mathcal{X}}$. For ease of notation we denote the subsequence again by $(\xi_n)_n$. Since $\widetilde{I}$ is lower-semicontinuous and $\xi \mapsto \Lambda(V,\xi)$ is continuous we have that

$$\lambda = \limsup_{n\to\infty} \left\{ \Lambda(V,\xi_n) - \widetilde{I}(\xi_n) \right\} \le \Lambda(V,\xi) - \widetilde{I}(\xi)\,,$$

from which we can conclude that ξ is a maximizer. We next claim that the maximizer has only one component. To see that assume that the maximizer ξ has at least two components, denoted by $\widetilde{\alpha}_1$ and $\widetilde{\alpha}_2$. Consider representants α_1 and α_2 such that $\int V(x,y)\alpha_1(dx)\,\alpha_2(dy) > 0$. This is possible since V is non-zero. We then define $\alpha_3 = \alpha_1 + \alpha_2$ and

$$\widetilde{\xi} = \{\alpha_3, \xi \setminus \{\alpha_1,\alpha_2\}\}\,.$$

Then,

$$\begin{aligned} \Lambda(V,\alpha_3) &= \int V(x,y)\alpha_1(dx)\,\alpha_1(dy) + \int V(x,y)\alpha_2(dx)\,\alpha_2(dy) + 2\int V(x,y)\alpha_1(dx)\,\alpha_2(dy) \\ &> \Lambda(V,\alpha_1) + \Lambda(V,\alpha_2)\,, \end{aligned}$$

by the choice of α_1 and α_2. By the sub-additivity proven in Lemma 2.13 we obtain that

$$I(\alpha_3) \leq I(\alpha_1) + I(\alpha_2)\,.$$

From these two observation it follows that

$$\Lambda(V,\xi) - \widetilde{I}(\xi) < \Lambda(V,\tilde{\xi}) - \widetilde{I}(\tilde{\xi})\,.$$

This however is a contradiction, thus proving that the maximizer ξ can only have one component. We next show that any maximizer is a probability measure. To that end assume that β is a sub-probability measure whose orbit maximizes the variational formula in (2.3.6). Define $p = \beta(\mathbb{Z}^d) \in (0,1)$ and consider the probability measure $\mu = \frac{\beta}{p}$ (it will follow from the arguments further below that $p > 0$). Then, using that V is a function of two variables and the definition of $\Lambda(V,\mu)$ in (2.1.1) shows that

$$\Lambda(V,\mu) - I(\mu) = \frac{1}{p}\Big(\frac{\Lambda(V,\beta)}{p} - I(\beta)\Big) \geq \frac{1}{p}(\Lambda(V,\beta) - I(\beta)) > \Lambda(V,\beta) - I(\beta)\,,$$

contradicting the assumption that β maximizes (2.3.6). Here we used that the supremum must be positive. Indeed, define

$$\tau_1 = \inf\{t \geq 0:\ X_t \neq 0\}\,,$$

which is the first jump time of the random walk X and has an exponential distribution with parameter $\sum_{y\in\mathbb{Z}^d} a_{0,y}$. It then follows that

$$Z_t \geq \mathbb{E}_0\Big[\exp\Big(t\int V(x,y)L_t(dx)L_t(dy)\Big)\mathbb{1}_{\{\tau_1>t\}}\Big] \geq \exp\Bigg(t\Big(V(0,0) - \sum_{y\in\mathbb{Z}^d} a_{0,y}\Big)\Bigg)\,,$$

and the claim follows by the assumption that $V(0,0) > \sum_{y\in\mathbb{Z}^d} a_{0,y}$. The above analysis implies the following result:

Theorem 2.23. *Consider a function $V : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}_+$ satisfying the assumptions stated in this section, then*

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}_0\Big[\exp\Big(\frac{1}{t}\int_0^t ds\int_0^t du\,V(X_s - X_u)\Big)\Big] = \sup_{\mu\in\mathcal{M}_1(\mathbb{Z}^d)}\Big\{\Lambda(V,\mu) - I(\mu)\Big\}\,.$$

The above variational formula has at least one maximizer. Moreover, if μ is a maximizer, then for any $x \in \mathbb{Z}^d$, the shifted measure $\mu_x := \mu * \delta_x$ is a maximizer as well.

Remark 2.24. It would be interesting to know under which conditions the above variational formula has a unique (unique modulo translations in space) maximizer. This, however, seems to be a delicate problem. Even in the case in which $\mathbb{Z}^d$ is replaced by $\mathbb{R}^d$ this is not a trivial issue, see [20]. We do not pursue this question here.

Bibliography

[1] E. Bates and S. Chatterjee. The endpoint distribution of directed polymers. *Ann. Probab.*, 48(2):817–871, 2020.

[2] E. Bolthausen. Localization of a two-dimensional random walk with an attractive path interaction. *Ann. Probab.*, 22(2):875–918, 1994.

[3] E. Bolthausen, W. König, and C. Mukherjee. Mean-field interaction of Brownian occupation measures II: A rigorous construction of the Pekar process. *Comm. Pure Appl. Math.*, 70(8):1598–1629, 2017.

[4] E. Bolthausen and U. Schmock. On self-attracting d-dimensional random walks. *Ann. Probab.*, 25(2):531–572, 1997.

[5] Y. Bröker and C. Mukherjee. Localization of the Gaussian multiplicative chaos in the Wiener space and the stochastic heat equation in strong disorder. *Ann. Appl. Probab.*, 29(6):3745–3785, 2019.

[6] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[7] F. den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[8] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.*, 28:1–47; ibid. 28 (1975), 279–301, 1975.

[9] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.*, 29(4):389–461, 1976.

[10] M. D. Donsker and S. R. S. Varadhan. On the number of distinct sites visited by a random walk. *Comm. Pure Appl. Math.*, 32(6):721–747, 1979.

[11] M. D. Donsker and S. R. S. Varadhan. Asymptotics for the polaron. *Comm. Pure Appl. Math.*, 35(4):505–528, 1983.

[12] D. Erhard and J. Poisat. Strong large deviation principles for pair empirical measures of random walks in the Mukherjee-Varadhan topology. *ArXiv*, 2023. `https://arxiv.org/abs/2304.00814`.

[13] Richard P. Feynman. *Statistical mechanics*. Advanced Book Classics. Perseus Books, Advanced Book Program, Reading, MA, 1998. A set of lectures, Reprint of the 1972 original.

[14] J. Gärtner and S.A. Molchanov. Parabolic problems for the anderson model ii. second-order asymptotics and structure of high peaks. *Probability Theory and Related Fields*, 111:17–55, 1998.

[15] W. König and C. Mukherjee. Mean-field interaction of Brownian occupation measures, I: Uniform tube property of the Coulomb functional. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):2214–2228, 2017.

[16] Wolfgang König. *The Parabolic Anderson Model: Random Walk in Random Potential*. Pathways in Mathematics. Birkhäuser, 2016.

[17] Wolfgang König. *Große Abweichungen: Techniken und Anwendungen*. Springer, 2020.

[18] Wolfgang König and Chiranjib Mukherjee. Mean-field interaction of Brownian occupation measures, I: Uniform tube property of the Coulomb functional. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):2214–2228, 2017.

[19] Yves Le Jan, Michael B. Marcus, and Jay Rosen. Intersection local times, loop soups and permanental wick powers. *Mem. Amer. Math. Soc.*, 247(1171):v+78, 2017.

[20] E. H. Lieb. Existence and uniqueness of the minimizing solution of choquard's nonlinear equation. *Stud.Appl. Math.*, 57(2):93–105, 1977.

[21] C. Mukherjee and S. R. S. Varadhan. Brownian occupation measures, compactness and large deviations. *Ann. Probab.*, 44(6):3934–3964, 2016.

[22] Herbert Spohn. Effective mass of the polaron: a functional integral approach. *Ann. Physics*, 175(2):278–318, 1987.

[23] S. R. S. Varadhan. The role of topology in large deviations. *Expo. Math.*, 36(3-4):362–368, 2018.