



**HAL**  
open science

# Strong large deviation principles for pair empirical measures of random walks in the Mukherjee-Varadhan topology

Dirk Erhard, Julien Poisat

► **To cite this version:**

Dirk Erhard, Julien Poisat. Strong large deviation principles for pair empirical measures of random walks in the Mukherjee-Varadhan topology. 2024. hal-04054395v2

**HAL Id: hal-04054395**

**<https://hal.science/hal-04054395v2>**

Preprint submitted on 11 Apr 2024

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

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

# STRONG LARGE DEVIATION PRINCIPLES FOR PAIR EMPIRICAL MEASURES OF RANDOM WALKS IN THE MUKHERJEE-VARADHAN TOPOLOGY

DIRK ERHARD AND JULIEN POISAT

ABSTRACT. In this paper we introduce a topology under which the pair empirical measure of a large class of random walks satisfies a strong Large Deviation principle. The definition of the topology is inspired by the recent article by Mukherjee and Varadhan [16]. This topology is natural for translation-invariant problems such as the downward deviations of the volume of a Wiener sausage or simple random walk, known as the Swiss cheese model [18]. We also adapt our result to some rescaled random walks and provide a contraction principle to the single empirical measure despite a lack of continuity from the projection map, using the notion of diagonal tightness.

## CONTENTS

1. Introduction	2
2. Topology on the space of probability measures modulo shifts	3
2.1. Vague and weak convergence	4
2.2. Widely separated sequences	4
2.3. Totally disintegrating sequences	6
2.4. Compactification	7
3. Large Deviation principles	11
4. Lower semi-continuity of the rate function	12
5. Lower bound	14
6. Upper bound	15
7. Adaptation to rescaled random walks	18
8. From the pair empirical measure to the empirical measure	19
8.1. Additional observations	26
Appendix A. On the Mukherjee-Varadhan topology	27
Acknowledgements	28
References	28

---

*Date:* April 11, 2024.

*2020 Mathematics Subject Classification.* 60F10 ; 60G37 ; 60J05 ; 54D35.

*Key words and phrases.* Large Deviation principles, empirical measure, occupation measure, random walk, compactification.

(D. Erhard) *Instituto de Matemática - Universidade Federal da Bahia*

(J. Poisat) *Université Paris-Dauphine, CNRS, UMR 7534, CEREMADE, PSL Research University, 75016 Paris, France .*

## 1. INTRODUCTION

Let  $X = (X_n)_{n \in \mathbb{N}_0}$  be a Markov chain on a Polish space  $\Sigma$  with a Feller transition probability  $\pi$  which we assume to have a density  $p(x, y)$  with respect to a reference measure  $\lambda(dy)$ . In the seventies, Donsker and Varadhan [9, (II)] showed that the empirical measure defined by

$$(1.1) \quad L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

satisfies a Large Deviation principle in the space of probability measures equipped with the weak topology, when  $\Sigma$  is *compact* and  $p(x, y)$  is uniformly bounded from above and below. If the state space  $\Sigma$  is *not compact* (e.g.  $\Sigma = \mathbb{R}^d$  equipped with the Euclidian distance), the upper bound holds for *compact* sets rather than *closed* sets, while the lower bound still holds for any open set under the following assumption: for all  $x \in \Sigma$ , for all measurable sets  $A \subset \Sigma$  such that  $\lambda(A) > 0$ , there exists  $k \in \mathbb{N}$  such that  $\pi^k(x, A) > 0$ , see [10, Corollary 3.4 and Equation (4.1)]. In some cases, this *weak* Large Deviation principle may be upgraded to a standard (or *strong*) one, recovering the Large Deviation upper bound for all closed sets. However such cases (see e.g. the uniformity assumption in [6, Corollary 6.5.10]) fail to include many natural examples. In applications, the lack of compactness may be dealt with by adding a confining drift to the Markov chain (or diffusion) [12] or folding it on a large torus [2, 4, 11]. Quite recently, Mukherjee and Varadhan [16] proposed a new approach in which they embed the space of probability measures on  $\mathbb{R}^d$  into a larger space equipped with a certain topology that makes it a *compact* metric space. This approach is actually very much in spirit of the concept of concentration-compactness introduced by P. L. Lions in the 80's, see [15], which cures the lack of compactness coming from the action of a noncompact group action like for instance rescaling or translation. Under this new topology, Mukherjee and Varadhan were able to prove a *strong* Large Deviation principle for the empirical measure of Brownian motion [16, Theorem 4.1], which was then successfully applied to the so-called polaron problem [3, 14, 16]. Recently, the compactification of measures has also proven fruitful in the context of directed polymers, see [1, 5]. The reader may refer to [6, 7, 8, 19] for an account on Large Deviation theory and to [20] for the role of topology in this theory.

In this paper we adapt and extend the work of Mukherjee and Varadhan in order to prove a *strong* Large Deviation principle (LDP) for the *pair* empirical measure of the Markov chain  $(X_n)_{n \in \mathbb{N}_0}$ , defined by

$$(1.2) \quad L_n^{(2)} = \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, X_{i+1})},$$

in the case  $\Sigma = \mathbb{R}^d$ . Our work is motivated by the application of this LDP to the so-called *Swiss cheese* problem, that is the (downward) Large Deviations for the volume of a Wiener sausage in  $\mathbb{R}^d$  ( $d \geq 2$ ), by van den Berg, Bolthausen and den Hollander [18]. In that paper (as well as in Phetpradap's Ph.D thesis [17] for the discrete random walk counterpart a few years later) the authors used the aforementioned folding procedure on a large torus to deal with the lack of compactness of the state space. In a companion paper [13], we apply the strong LDP for the pair empirical measure to the Swiss cheese

problem in order to obtain the so-called *tube property*. The latter essentially means that, conditioned to the downward deviations of the random walk volume, the empirical measure of a certain random walk skeleton introduced in [18] is close to one of the minimizers of the variational problem in the corresponding rate function.

The present paper is organized as follows. In Section 2 we introduce the relevant notation and topology. Although there is a lot in common with [16], let us stress that this is *not* the Mukherjee-Varadhan topology applied to the product space  $\mathbb{R}^d \times \mathbb{R}^d$  instead of  $\mathbb{R}^d$ , see also Remark 2.1 below. Our main result, that is the Large Deviation principle for the pair empirical measure of a large class of random walks, in this new topology, is stated in Theorem 3.1 of Section 3. Lower semi-continuity of the rate function is proven in Section 4. The lower and upper bounds of the LDP are proven in Sections 5 and 6, respectively. We shall use therein the well-known fact that  $(X_n, X_{n+1})_{n \in \mathbb{N}_0}$  is itself a Markov chain. In Section 7 we adapt our result to the case of certain rescaled random walks. Finally, in Section 8 we explain how to use Theorem 3.1 to obtain a Large Deviation principle for the (single) empirical measure. This part is non-trivial since, due to a lack of continuity from the projection map, the standard contraction principle is not applicable. We use the new notion of *diagonal tightness* to overcome this difficulty. In order to make the paper more self-contained, we include a condensed presentation of the Mukherjee-Varadhan topology in Appendix A.

## 2. TOPOLOGY ON THE SPACE OF PROBABILITY MEASURES MODULO SHIFTS

Throughout the paper, we equip  $\mathbb{R}^d$  with any of its equivalent norms, further denoted by  $|\cdot|$ , while  $\mathbb{R}^d \times \mathbb{R}^d$  is equipped with the product norm  $\|(x, y)\| = |x| \vee |y| = \max(|x|, |y|)$ . Let  $\mathcal{M}_1^{(2)} := \mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$  be the space of probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$  and  $\mathcal{M}_{\leq 1}^{(2)} := \mathcal{M}_{\leq 1}(\mathbb{R}^d \times \mathbb{R}^d)$  be the space of sub-probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$ . We consider the action of the *diagonal* shifts  $\theta_{x,x}$  for  $x \in \mathbb{R}^d$ , defined by:

$$(2.1) \quad \int_{\mathbb{R}^d \times \mathbb{R}^d} f(u, v) (\theta_{x,x} \nu) (du, dv) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(u+x, v+x) \nu(du, dv)$$

for all continuous bounded functions  $f: \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$  and  $\nu \in \mathcal{M}_{\leq 1}^{(2)}$ . We shall denote by  $\widetilde{\mathcal{M}}_1^{(2)}$  (resp.  $\widetilde{\mathcal{M}}_{\leq 1}^{(2)}$ ) the space of equivalence classes of  $\mathcal{M}_1^{(2)}$  (resp.  $\mathcal{M}_{\leq 1}^{(2)}$ ) under the collection of shifts  $\theta_{x,x}$ . For  $k \geq 2$ , we define  $\mathcal{F}_k$  as the space of continuous functions  $f: (\mathbb{R}^d)^k \mapsto \mathbb{R}$  that are translation invariant, i.e.

$$(2.2) \quad f(x_1 + x, \dots, x_k + x) = f(x_1, \dots, x_k), \quad \forall x, x_1, \dots, x_k \in \mathbb{R}^d,$$

and *vanishing at infinity*, in the sense that

$$(2.3) \quad \lim_{\max_{i \neq j} |x_i - x_j| \rightarrow \infty} f(x_1, \dots, x_k) = 0.$$

Let us define  $\mathcal{F}_k^{(2)} = \mathcal{F}_{2k}$  for  $k \geq 1$ . When  $f \in \mathcal{F}_k^{(2)}$  and  $\alpha \in \mathcal{M}_{\leq 1}^{(2)}$ , we write

$$(2.4) \quad \Lambda(f, \alpha) := \int f(u_1, v_1, \dots, u_k, v_k) \prod_{1 \leq i \leq k} \alpha(du_i, dv_i),$$

which only depends on the orbit  $\tilde{\alpha}$ .

**Remark 2.1.** Note that the space  $\widetilde{\mathcal{M}}_1^{(2)}$  defined here is different from  $\widetilde{\mathcal{M}}_1(\mathbb{R}^{2d})$  defined in [16]. Indeed, in [16] the shifts are with respect to all directions in  $\mathbb{R}^{2d}$  whereas here they are only with respect to all directions of the form  $(x, x) \in \mathbb{R}^{2d}$ . This is very natural in view of our application to the pair empirical measure.

**Remark 2.2.** There is another natural choice for the set of test functions. We could say that  $f$  (with an even number of arguments) is vanishing at infinity if

$$(2.5) \quad \lim_{\max_{i \neq j} \|(u_i, v_i) - (u_j, v_j)\| \rightarrow \infty} f(u_1, v_1, \dots, u_k, v_k) = 0.$$

Note that this is a larger set of test functions. Indeed, if we write  $(x_1, x_2, \dots, x_{2k}) = (u_1, v_1, \dots, u_k, v_k)$  then

$$(2.6) \quad \max_{i \neq j} |x_i - x_j| \geq \max_{i \neq j} |u_i - u_j| \vee \max_{i \neq j} |v_i - v_j| = \max_{i \neq j} \|(u_i, v_i) - (u_j, v_j)\|.$$

This choice would then lead to the definition of an a priori stronger topology.

**2.1. Vague and weak convergence.** In this section we give a slight generalisation of [16, Lemma 2.2] which will be useful for later purposes.

**Lemma 2.3.** Consider a sequence  $(\mu_n)_{n \in \mathbb{N}_0}$  of sub-probability measures in  $\mathbb{R}^d$  that converges vaguely to some sub-probability measure  $\alpha$ . Then we can write  $\mu_n = \alpha_n + \beta_n$  for every  $n \in \mathbb{N}_0$ , where  $(\alpha_n)_{n \in \mathbb{N}_0}$  converges weakly to  $\alpha$  and  $(\beta_n)_{n \in \mathbb{N}_0}$  converges vaguely to zero. Moreover,  $\alpha_n$  and  $\beta_n$  can be chosen such that they have disjoint supports.

*Proof of Lemma 2.3.* The only part that was not explicitly mentioned in [16] is that  $\alpha_n$  and  $\beta_n$  can be chosen to have disjoint supports. However, this is a direct consequence of the proof in [16]. Indeed, adopting their notation they defined  $\alpha_n$  as the restriction of  $\mu_n$  to  $B(0, R_n)$  and  $\beta_n$  as the restriction of  $\mu_n$  to  $B(0, R_n)^c$ , where  $\|\cdot\|_2$  denotes the Euclidean norm in  $\mathbb{R}^d$  and  $(R_n)_{n \in \mathbb{N}_0}$  is an increasing sequence converging to infinity satisfying

$$(2.7) \quad \mu_n(B(0, R_n)) \leq \alpha(\mathbb{R}^d) + \frac{1}{R_n}.$$

The claim therefore follows. ■

**2.2. Widely separated sequences.** We say that two sequences  $(\alpha_n)$  and  $(\beta_n)$  in  $\mathcal{M}_{\leq 1}^{(2)}$  are *widely separated* if, for some positive function  $V$  in  $\mathcal{F}_2^{(2)}$ ,

$$(2.8) \quad \lim_{n \rightarrow \infty} \int_{(\mathbb{R}^d)^4} V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) = 0.$$

**Remark 2.4.** There exist  $(\alpha_n)$  and  $(\beta_n)$  that are widely separated for our choice of test functions but not widely separated for the larger set of test functions mentioned in Remark 2.2. Indeed, let  $\alpha_n$  be a Dirac mass at  $(u(n), v(n)) \in \mathbb{R}^d \times \mathbb{R}^d$ , where  $|u(n) - v(n)| \rightarrow \infty$ . It is clear that  $(\alpha_n)$  is widely separated (in our sense) from any other sequence (including itself). Now, consider the bounded continuous function

$$(2.9) \quad V(u_1, v_1, u_2, v_2) = \frac{1}{(1 + |u_1 - u_2|)(1 + |v_1 - v_2|)}.$$

The function  $V$  is not vanishing at infinity in our sense, since  $V(u_1, v_1, u_2, v_2) = V(u_1 + x, v_1, u_2 + x, v_2)$  for every  $x \in \mathbb{R}^d$ , but it is vanishing at infinity in the less restrictive sense of Remark 2.2 and

$$(2.10) \quad \int V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \alpha_n(du_2, dv_2) = 1$$

does not converge to zero.

The following lemma, which mimicks [16, Lemma 2.4], lists the most important properties of widely separated sequences of sub-probability measures.

**Lemma 2.5.** *Let  $(\alpha_n)$  and  $(\beta_n)$  be two widely separated sequences in  $\mathcal{M}_{\leq 1}^{(2)}$ . Then,*

(1) *For every  $W \in \mathcal{F}_2^{(2)}$ ,*

$$(2.11) \quad \lim_{n \rightarrow \infty} \int W(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) = 0.$$

(2) *For every  $k \geq 2$  and every  $f \in \mathcal{F}_k^{(2)}$ ,*

$$(2.12) \quad \lim_{n \rightarrow \infty} |\Lambda(f, \alpha_n + \beta_n) - \Lambda(f, \alpha_n) - \Lambda(f, \beta_n)| = 0.$$

*Proof of Lemma 2.5.* Let  $W \in \mathcal{F}_2^{(2)}$  be arbitrary and let  $V \in \mathcal{F}_2^{(2)}$  be the positive function from (2.8). Then, since  $V(0, v_1, u_2, v_2)$  is bounded from below by a positive constant on compact sets, for every  $\varepsilon > 0$  there exists a constant  $C_\varepsilon$  such that for any  $v_1, u_2, v_2 \in \mathbb{R}^d$ ,

$$(2.13) \quad |W(0, v_1, u_2, v_2)| \leq C_\varepsilon V(0, v_1, u_2, v_2) + \varepsilon.$$

Thus

$$(2.14) \quad \begin{aligned} & \limsup_{n \rightarrow \infty} \int |W(u_1, v_1, u_2, v_2)| \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) \\ & \leq C_\varepsilon \limsup_{n \rightarrow \infty} \int V(0, v_1 - u_1, u_2 - u_1, v_2 - u_1) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) + \varepsilon \\ & = \varepsilon, \end{aligned}$$

and (1) follows. To see that (2) holds as well, we first note that the case  $k = 2$  is a direct consequence of the first part of the lemma. The case  $k \geq 3$  follows easily: w.l.o.g, any cross-term in the expansion of  $\Lambda(f, \alpha_n + \beta_n)$  may be written as

$$(2.15) \quad \int f(u_1, v_1, u_2, v_2, \dots, u_k, v_k) \alpha_n(u_1, v_1) \beta_n(u_2, v_2) \prod_{3 \leq i \leq k} \gamma_{n,i}(u_i, v_i),$$

where  $\gamma_{n,i}$  is either  $\alpha_n$  or  $\beta_n$ . Using translation invariance and repeating the argument used at the beginning of the proof, we see that for every  $\varepsilon > 0$ , there exists  $C_\varepsilon > 0$  such that for all  $v_1, u_2, v_2, \dots, u_k, v_k \in \mathbb{R}^d$ ,

$$(2.16) \quad |f(0, v_1, u_2, v_2, \dots, u_k, v_k)| \leq C_\varepsilon V(0, v_1, u_2, v_2) + \varepsilon,$$

which allows to conclude. ■

Here is a sufficient condition for two sequences of measures to be widely separated.

**Lemma 2.6.** *Let  $(\alpha_n)$  and  $(\beta_n)$  be two sequences in  $\mathcal{M}_{\leq 1}^{(2)}$ . If  $(\alpha_n)$  is tight and  $(\beta_n)$  converges vaguely to zero, then they are widely separated.*

*Proof of Lemma 2.6.* Let  $V \in \mathcal{F}_2^{(2)}$  and  $\varepsilon > 0$ . By tightness of  $(\alpha_n)$  and boundedness of  $V$ , there exists  $M > 0$  such that for all  $n \geq 1$ ,

$$(2.17) \quad \int V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) \leq \int_{|u_1|, |v_1| \leq M} (\dots) + \varepsilon \|V\|_\infty.$$

Here,  $(\dots)$  stands of course for  $V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2)$ . We further split the integral on the right-hand side as

$$(2.18) \quad \int_{\substack{|u_1|, |v_1| \leq M \\ |u_2|, |v_2| \leq 2M}} (\dots) + \int_{\substack{|u_1|, |v_1| \leq M \\ |u_2| \vee |v_2| > 2M}} (\dots).$$

We claim that we can make the second term smaller than  $\varepsilon$  by choosing  $M$  even larger if necessary. Indeed, since  $|u_2 - u_1| \vee |v_2 - v_1| \geq M$  on the domain of integration and  $V$  is vanishing at infinity the claim follows. As for the first term, it goes to zero as  $n$  goes to infinity since  $V$  is bounded and  $(\beta_n)$  converges vaguely to zero.  $\blacksquare$

**2.3. Totally disintegrating sequences.** We say that a sequence  $(\mu_n)$  in  $\mathcal{M}_{\leq 1}^{(2)}$  is *totally disintegrating* if, for any  $r > 0$ ,

$$(2.19) \quad \lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} \mu_n(B((x, x), r)) = 0.$$

From now on and unless stated otherwise the balls are defined with respect to the  $\ell_\infty$ -norm. Following [16, Lemma 2.3], we obtain the following result.

**Lemma 2.7.** *The sequence  $(\mu_n)$  in  $\mathcal{M}_{\leq 1}^{(2)}$  is totally disintegrating iff one of the following equivalent statements holds:*

- (1) *There exists a positive  $V \in \mathcal{F}_2^{(2)}$  such that  $\lim_{n \rightarrow \infty} \Lambda(V, \mu_n) = 0$ .*
- (2) *For any  $V \in \mathcal{F}_2^{(2)}$ ,  $\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} \int V(0, x, u, v) \mu_n(du, dv) = 0$ .*
- (3) *For any  $V \in \mathcal{F}_2^{(2)}$ ,  $\lim_{n \rightarrow \infty} \Lambda(V, \mu_n) = 0$ .*

*Proof of Lemma 2.7.* We prove that (1)  $\Rightarrow$  (2.19)  $\Rightarrow$  (2)  $\Rightarrow$  (3). Clearly, (3) implies (1).

(i) Let us prove that (1) implies (2.19). Letting

$$(2.20) \quad \delta := \min_{a, b, c \in B(0, 2r)} V(0, a, b, c) > 0,$$

we get

$$(2.21) \quad \begin{aligned} & \int V(u_1, v_1, u_2, v_2) \mu_n(du_1, dv_1) \mu_n(du_2, dv_2) \\ &= \int V(0, v_1 - u_1, u_2 - u_1, v_2 - u_1) \mu_n(du_1, dv_1) \mu_n(du_2, dv_2) \\ &\geq \delta \int_{v_1, u_2, v_2 \in B(u_1, 2r)} \mu_n(du_1, dv_1) \mu_n(du_2, dv_2) \\ &\geq \delta \sup_{x \in \mathbb{R}^d} \mu_n(B((x, x), r))^2. \end{aligned}$$

(ii) From (2.19) to (2). Let  $x \in \mathbb{R}^d$ . Fix  $M > 0$  and write

$$(2.22) \quad \begin{aligned} & \int |V(0, x, u, v)| \mu_n(du, dv) \\ &= \int_{B((x,x),M)} |V(0, x, u, v)| \mu_n(du, dv) + \int_{B((x,x),M)^c} |V(0, x, u, v)| \mu_n(du, dv). \end{aligned}$$

The first term on the right-hand side goes to zero uniformly in  $x$  as  $n$  tends to infinity by the boundedness of  $V$  and by (2.19), while the second term goes to zero uniformly in  $x$  as  $M$  tends to infinity by the fact that  $V$  vanishes at infinity.

(iii) To go from (2) to (3), write

$$(2.23) \quad \begin{aligned} \Lambda(V, \mu_n) &\leq \sup_{x, y \in \mathbb{R}^d} \int V(0, y - x, u - x, v - x) \mu_n(du, dv) \\ &\leq \sup_{y \in \mathbb{R}^d} \int V(0, y, u, v) \mu_n(du, dv), \end{aligned}$$

which goes to zero as  $n \rightarrow \infty$ , by assumption. Here, we used that  $\mu_n$  is a sub-probability measure to obtain the first inequality.  $\blacksquare$

As an immediate corollary we obtain the following:

**Corollary 2.8.** *If the sequence  $(\mu_n)$  is totally disintegrating then, for any  $k \geq 2$  and any  $V \in \mathcal{F}_k^{(2)}$ ,*

$$(2.24) \quad \lim_{n \rightarrow \infty} \Lambda(V, \mu_n) = 0.$$

*Proof of Corollary 2.8.* Apply Item (3) in Lemma 2.7 to

$$(2.25) \quad W(u_1, v_1, u_2, v_2) := \sup_{u_3, v_3, \dots, u_k, v_k \in \mathbb{R}^d} |V(u_1, v_1, \dots, u_k, v_k)|.$$

$\blacksquare$

**2.4. Compactification.** Let us define

$$(2.26) \quad \mathcal{F}^{(2)} := \bigcup_{k \geq 2} \mathcal{F}_k^{(2)},$$

for which there exists a countable dense set (under the uniform metric) denoted by

$$(2.27) \quad \{f_r(u_1, v_1, \dots, u_{k_r}, v_{k_r}) : r \in \mathbb{N}\},$$

(same arguments as in [16, Section 2.2]). We define

$$(2.28) \quad \tilde{\mathcal{X}}^{(2)} := \left\{ \xi = \{\tilde{\alpha}_i\}_{i \in I} : \alpha_i \in \mathcal{M}_{\leq 1}^{(2)}, \sum_{i \in I} \alpha_i(\mathbb{R}^d \times \mathbb{R}^d) \leq 1 \right\},$$

where  $I$  may be empty, finite or countable. Recall (2.4) and for any  $\xi_1, \xi_2 \in \tilde{\mathcal{X}}^{(2)}$ , define

$$(2.29) \quad \mathbf{D}^{(2)}(\xi_1, \xi_2) := \sum_{r \geq 1} \frac{1}{2^r} \frac{1}{1 + \|f_r\|_\infty} \left| \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f_r, \tilde{\alpha}) - \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f_r, \tilde{\alpha}) \right|.$$

**Proposition 2.9.**  $\mathbf{D}^{(2)}$  is a metric on  $\tilde{\mathcal{X}}^{(2)}$ .



*Proof of Proposition 2.9.* As in [16, Theorem 3.1], the only non-trivial part is to show that  $\mathbf{D}^{(2)}(\xi_1, \xi_2) = 0$  implies  $\xi_1 = \xi_2$ . We follow the three-step proof from that reference.

**Step 1.** We give some level of details to show how the proof is adapted to our case. If  $\mathbf{D}^{(2)}(\xi_1, \xi_2) = 0$  then for all  $k \geq 2$  and  $f \in \mathcal{F}_k^{(2)}$ ,

$$(2.30) \quad \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f, \alpha) = \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f, \alpha).$$

We deduce therefore that for every integer  $r \geq 1$ ,

$$(2.31) \quad \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f, \alpha)^r = \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f, \alpha)^r.$$

Indeed, define the function (for  $r = 2$ )

$$(2.32) \quad \begin{aligned} &g_N(u_1, v_1, \dots, u_k, v_k, u_{k+1}, v_{k+1}, \dots, u_{2k}, v_{2k}) \\ &:= f(u_1, v_1, \dots, u_k, v_k) f(u_{k+1}, v_{k+1}, \dots, u_{2k}, v_{2k}) \varphi\left(\frac{u_{k+1} - u_1}{N}\right), \end{aligned}$$

where  $0 \leq \varphi \leq 1$  is equal to 1 inside a ball of radius 1 and is truncated smoothly to 0 outside a ball of radius 2. Then,  $g_N \in \mathcal{F}_{4k}$  and converges pointwise to  $f \otimes f$  as  $N \rightarrow \infty$ . Hence, using the fact that  $f$  is bounded and dominated convergence

$$(2.33) \quad \lim_{N \rightarrow \infty} \Lambda(g_N, \alpha) = \Lambda(f, \alpha)^2.$$

The general case follows the same idea.

**Step 2.** From there we prove that to every orbit  $\tilde{\alpha}_1$  in  $\xi_1$  we can match a single orbit  $\tilde{\alpha}_2$  in  $\xi_2$  such that  $\Lambda(f, \alpha_1) = \Lambda(f, \alpha_2)$  for all  $f \in \mathcal{F}_k^{(2)}$  and  $k \geq 2$ . This part of the argument is exactly the same as in [16]. We do not reproduce it here, for conciseness.

**Step 3.** This step is also an adaptation of [16], so we only sketch the arguments. We want to recover the orbit of  $\alpha \in \mathcal{M}_{\leq 1}^{(2)}$  from the value of  $\Lambda(f, \alpha)$  for  $f \in \mathcal{F}_k^{(2)}$ . Adapting [16], from these values we get those of

$$(2.34) \quad \prod_{j=1}^k \phi(s_j, t_j), \quad \text{where} \quad \phi(s, t) := \int e^{i\langle (s,t), (u,v) \rangle} \alpha(du, dv), \quad s, t \in \mathbb{R}^d,$$

provided  $\sum(s_j + t_j) = 0$ . Suppose now that

$$(2.35) \quad \forall k \geq 1, \quad \prod_{j=1}^k \phi(s_j, t_j) = \prod_{j=1}^k \psi(s_j, t_j), \quad \text{whenever} \quad \sum(s_j + t_j) = 0,$$

where  $\psi$  is another characteristic function. Following [16], we obtain that  $|\phi(s, t)| = |\psi(s, t)|$  and write  $\phi(s, t) = \psi(s, t)\chi(s, t)$  whenever  $|\phi(s, t)| = |\psi(s, t)| \neq 0$ . As soon as  $\sum(s_j + t_j) = \sigma + \tau \in \mathbb{R}^d$ , we have  $\prod_{j=1}^k \chi(s_j, t_j) = \chi(\sigma, \tau)$ , provided that the  $s_j$ 's and  $t_j$ 's are such that  $|\psi(s_j, t_j)| = |\phi(s_j, t_j)| \neq 0$  for all  $j$ , and  $\chi(\sigma, \tau) \neq 0$ . In particular,

$$(2.36) \quad \chi(s_1 + s_2, t_1 + t_2) = \chi(s_1, t_1)\chi(s_2, t_2),$$

hence as in [16] we can show that  $\chi(s, t) = e^{i\langle (a_1, s) + (a_2, t) \rangle}$  for some  $a_1, a_2 \in \mathbb{R}^d$ . The fact that actually  $\chi(s_1 + s_2 + C, t_1 + t_2 - C) = \chi(s_1, t_1)\chi(s_2, t_2)$  for all  $C \in \mathbb{R}^d$  entails that

$a_1 = a_2$ . This means that  $\alpha$  is determined up to shifts by some  $(a, a) \in \mathbb{R}^{2d}$ , which ends the proof.  $\blacksquare$

**Proposition 2.10.** *The space  $\tilde{\mathcal{X}}^{(2)}$  equipped with  $\mathbf{D}^{(2)}$  is a compact metric space that contains  $\tilde{\mathcal{M}}_1^{(2)}$  as a dense subset.*

*Proof of Proposition 2.10.* This proof is an adaptation of the arguments used in [16], so we only sketch the arguments.

**Step 1.** Let us first show that  $\tilde{\mathcal{M}}_1^{(2)}$  is dense in  $\tilde{\mathcal{X}}^{(2)}$ . Let  $\xi = \{\tilde{\alpha}_i, i \in I\} \in \tilde{\mathcal{X}}^{(2)}$  and  $\varepsilon > 0$ . Pick a finite collection  $\{\tilde{\alpha}_i, 1 \leq i \leq k\}$  such that  $\sum_{i>k} p_i < \varepsilon$ , where  $p_i$  denotes the total mass of  $\tilde{\alpha}_i$ . Let  $\alpha_i$  be an arbitrary member of the orbit  $\tilde{\alpha}_i$  and for  $M > 0$ , let  $\lambda_M$  be the Gaussian law on  $\mathbb{R}^{2d}$  with zero mean and covariance matrix  $M \times \text{Id}$ . Pick any sequence  $(a_{i,M})_{1 \leq i \leq k}$  in  $(\mathbb{R}^d)^k$  such that  $\inf_{i \neq j} |a_{i,M} - a_{j,M}| \rightarrow \infty$  as  $M \rightarrow \infty$ . Finally, set

$$(2.37) \quad \mu_M := \sum_{1 \leq i \leq k} \alpha_i * \delta_{(a_{i,M}, a_{i,M})} + \left(1 - \sum_{1 \leq i \leq k} p_i\right) \lambda_M \in \mathcal{M}_1^{(2)}.$$

Since the mutual distances of the centers of masses of the measures in the first sum increase to infinity, i.e., they are widely separated, see (2.8) and all the mass of  $\lambda_M$  vanishes in the limit as  $M$  tends to infinity, i.e,  $\lambda_M$  is totally disintegrating, see (2.19), we get by Lemmas 2.5 and 2.7 that for all  $k \geq 1$  and  $f \in \mathcal{F}_k^{(2)}$ ,

$$(2.38) \quad \lim_{M \rightarrow \infty} \Lambda(f, \mu_M) = \sum_{1 \leq i \leq k} \Lambda(f, \tilde{\alpha}_i).$$

Since  $\varepsilon > 0$  may be chosen arbitrarily small this completes the first step.

**Step 2.** Let us show that for any sequence  $(\mu_n)$  in  $\mathcal{M}_1^{(2)}$ , there exists a subsequence along which  $(\tilde{\mu}_n)$  converges to some element of  $\tilde{\mathcal{X}}^{(2)}$ . Together with the first step this implies the result. We start with some preliminary considerations. We use the following *concentration function*:

$$(2.39) \quad q_\mu(r) := \sup_{x \in \mathbb{R}^d} \mu(B((x, x), r)), \quad r \geq 0, \quad \mu \in \mathcal{M}_1^{(2)}.$$

Let now  $(\mu_n)_{n \in \mathbb{N}}$  be a sequence in  $\mathcal{M}_{\leq 1}^{(2)}$ . Going over to a subsequence if necessary, we may define, by Helly's selection theorem,

$$(2.40) \quad q := \lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} q_{\mu_n}(r) \quad \text{and} \quad p := \lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^{2d}).$$

If  $q = 0$ , then it follows by Corollary 2.8 that  $\tilde{\mu}_n \rightarrow 0$  in  $\tilde{\mathcal{X}}^{(2)}$ . If on the other hand  $q > 0$ , then taking a suitable translation vector  $(a_n, a_n) \in \mathbb{R}^d \times \mathbb{R}^d$  we have that for some  $r > 0$  and all sufficiently large  $n$  the shifted measure  $\lambda_n = \mu_n * \delta_{(a_n, a_n)}$  satisfies

$$(2.41) \quad \lambda_n(B((0, 0), r)) \geq q/2.$$

Choosing a subsequence if needed, we may assume that  $(\lambda_n)$  converges vaguely to some  $\alpha \in \mathcal{M}_{\leq 1}^{(2)}$  and by Lemma 2.3 we may further write  $\lambda_n = \alpha_n + \beta_n$  with  $\alpha_n$  converging

weakly to  $\alpha$ ,  $\beta_n$  converging vaguely to zero and  $\alpha_n$  and  $\beta_n$  having disjoint supports. By Lemma 2.6\*, for every  $V \in \mathcal{F}_2^{(2)}$ ,

$$(2.42) \quad \lim_{n \rightarrow \infty} \int V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) = 0.$$

If additionally  $q = p$ , then no mass escapes to infinity and one can choose  $\beta_n$  to be zero. In that case it follows that  $\tilde{\mu}_n \rightarrow \alpha$  in  $\tilde{\mathcal{X}}^{(2)}$ , similarly to [16, Theorem 3.2]. To conclude the result we can now proceed in very much the same way as in [16]. Fix a sequence  $(\mu_n)_n$  in  $\mathcal{M}_1^{(2)}$ . If  $q \in \{0, 1\}$  then the above considerations imply the result. Otherwise, for some sequence  $(a_n, a_n) \in \mathbb{R}^d \times \mathbb{R}^d$ , at least along a subsequence, we have the decomposition  $\mu_n = \alpha_n + \beta_n$ , where

- (1)  $\alpha_n * \delta_{(a_n, a_n)} \rightarrow \alpha$  weakly as  $n \rightarrow \infty$ ;
- (2) For every  $V \in \mathcal{F}_2^{(2)}$  the integral

$$\int V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2)$$

converges to zero;

- (3)  $\lim_{n \rightarrow \infty} q_{\beta_n}(r) \leq \min\{1 - q/2, q\}$ ;
- (4)  $\alpha_n$  and  $\beta_n$  have disjoint support.

In the last item we used that mass  $q/2$  has been removed from  $\mu_n$ , see (2.41). If the above limit is not zero one can repeat the same procedure with  $\beta_n$ , i.e., for an appropriate shift of  $\beta_n$  one decomposes it as a sum of two measures where one converges weakly and the other converges vaguely to zero and such that additionally the above three items are satisfied. If this process terminates after some finite number of stages  $k \in \mathbb{N}$  we obtain the decomposition

$$(2.43) \quad \mu_n = \sum_{j=1}^k \alpha_{n,j} * \delta_{(a_{n,j}, a_{n,j})} + \beta_n,$$

such that  $\alpha_{n,j} * \delta_{(a_{n,j}, a_{n,j})}$  converges weakly,  $q_{\beta_n}(r) \rightarrow 0$  for every  $r$  and such that for every  $V \in \mathcal{F}_2^{(2)}$  and  $1 \leq i < j \leq k$ ,

$$(2.44) \quad \int V(u_1 + a_{n,i}, v_1 + a_{n,i}, u_2 + a_{n,j}, v_2 + a_{n,j}) \alpha_{n,i}(du_1, dv_1) \alpha_{n,j}(du_2, dv_2)$$

and

$$\int V(u_1 + a_{n,i}, v_1 + a_{n,i}, u_2, v_2) \alpha_{n,i}(du_1, dv_1) \beta_n(du_2, dv_2)$$

tend to zero as  $n \rightarrow \infty$ . If the process does not terminate after a finite number of stages, then one has a similar decomposition that goes by induction. We refer to the proof of [16, Theorem 3.2] for details. ■

---

\*The use of Lemma 2.6 corrects a minor gap in [16]. Actually, the analogous step in the proof of [16, Theorem 3.2, Step 2] does not seem to follow from Lemma 2.4, as it is claimed therein. Indeed, a sequence of sub-probability measures may converge vaguely to zero without disintegrating (think of a Dirac mass escaping to infinity).

### 3. LARGE DEVIATION PRINCIPLES

Let  $X = (X_i)_{i \in \mathbb{N}_0}$  be a Markov chain in  $\mathbb{R}^d$  starting from the origin and with a Feller transition kernel  $\pi$  satisfying the following assumptions:

- (1) **(Random walk)** There exists a function  $p: \mathbb{R}^d \mapsto \mathbb{R}^+$  and a reference measure  $\lambda$  such that  $\pi(x, A) = \int_A p(y-x)\lambda(dy)$  for all  $x \in \mathbb{R}^d$  and Borel sets  $A \subseteq \mathbb{R}^d$ .
- (2) **(Irreducibility)** For all Borel sets  $A \subseteq \mathbb{R}^d$  such that  $\lambda(A) > 0$ , there exists  $k \in \mathbb{N}$  such that  $\pi^k(x, A) = \int_A p^{*k}(y-x)\lambda(dy) > 0$ .
- (3) **(Tightness)** There exists a positive sequence  $(\rho_n)_{n \in \mathbb{N}}$  with  $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \rho_n = 0$  such that

$$(3.1) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \sup_{1 \leq i \leq n} |X_i| \geq \rho_n \right) = -\infty.$$

These assumptions include many natural examples such as the simple random walk on  $\mathbb{Z}^d$  (with the counting measure as reference measure) or the discretized Brownian motion  $(B_{i\varepsilon})_{i \in \mathbb{N}_0}$ , where  $\varepsilon > 0$  and  $B$  is Brownian motion (with Lebesgue measure as reference measure). Assumption (1) implies shift-invariance of the process, see Remark 3.2 below. Assumption (2) is used at the end of the proof of Proposition 5.1 when applying the standard Large Deviation lower bound in the usual weak topology [10, Corollary 3.4 and Equation (4.1)]. Assumption (3) is used during the proof of Lemma 6.2.

Let us denote by

$$(3.2) \quad L_n^{(2)} := \frac{1}{n} \sum_{i=1}^n \delta_{(X_{i-1}, X_i)} \in \mathcal{M}_1^{(2)}$$

the pair empirical measure associated to  $X$  and by  $\tilde{L}_n^{(2)}$  its orbit in  $\tilde{\mathcal{M}}_1^{(2)}$ . In the following,  $h(\cdot|\cdot)$  is the relative entropy between two measures, i.e.,

$$(3.3) \quad h(\mu|\nu) = \begin{cases} \int \log \left( \frac{d\mu}{d\nu} \right) d\mu & \text{if } \mu \ll \nu, \\ +\infty & \text{else.} \end{cases}$$

Note that when  $\mu$  is the zero or empty measure it makes sense to let  $h(\mu|\nu) = 0$ . Finally, when  $\nu \in \mathcal{M}_1(\mathbb{R}^d)$ , we write

$$(3.4) \quad (\nu \otimes \pi)(dx, dy) = \nu(dx)\pi(x, dy).$$

We may now state our main result:

**Theorem 3.1.** *As  $n \rightarrow \infty$ ,  $(\tilde{L}_n^{(2)})_{n \in \mathbb{N}}$  satisfies a strong Large Deviation principle on the compact metric space  $\tilde{\mathcal{X}}^{(2)}$  equipped with  $\mathbf{D}^{(2)}$ , with speed  $n$  and rate function  $\tilde{J}^{(2)}$ , where*

$$(3.5) \quad \tilde{J}^{(2)}(\xi) := \sum_{\alpha \in \xi} h(\alpha|\alpha_1 \otimes \pi),$$

if  $\alpha_1 = \alpha_2$  for all  $\alpha \in \xi$ , and  $\tilde{J}^{(2)}(\xi) := +\infty$  otherwise. Here we write, with a slight abuse of notation,  $\alpha \in \mathcal{M}_{\leq 1}^{(2)}$  for a representative of  $\alpha \in \xi$ , while  $\alpha_1$  and  $\alpha_2$  are the projections of the representative  $\alpha$  onto the first and the last  $d$  coordinates respectively.

**Remark 3.2.** *The rate function in (3.5) is well-defined due to the fact that the transition kernel  $\pi$  only depends on the difference of its two arguments. Indeed, this implies that inside the sum in (3.5) the choices of the representants do not matter.*

**Remark 3.3.** *Following the same idea, one should be able to prove a strong Large Deviation principle for the  $k$ -th order empirical measure ( $k \geq 2$ ) defined as*

$$(3.6) \quad L_n^{(k)} := \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, X_{i+1}, \dots, X_{i+k-1})},$$

see [6, Section 6.5.2] for the standard version of it. Naturally, the orbits of the sub-probability measures on  $\mathbb{R}^{kd}$  would be taken with respect to diagonal shifts of the form  $(x, x, \dots, x) \in \mathbb{R}^{kd}$ , where  $x \in \mathbb{R}^d$ . In analogy with the case  $k = 2$ , we expect the rate function to read

$$(3.7) \quad \xi \in \tilde{\mathcal{X}}^{(k)} \mapsto \tilde{J}^{(k)}(\xi) := \sum_{\alpha \in \xi} h(\alpha | \alpha_{1\dots k-1} \otimes \pi),$$

if  $\alpha_{1\dots k-1}$  (projection onto the  $k - 1$  first  $\mathbb{R}^d$ -valued coordinates) coincides with  $\alpha_{2\dots k}$  (projection onto the  $k - 1$  last  $\mathbb{R}^d$ -valued coordinates) for all  $\alpha \in \xi$ , and  $\tilde{J}^{(k)}(\xi) := +\infty$  otherwise. Note that we used a notation analogous to (3.4), namely:

$$(3.8) \quad \begin{aligned} (\alpha_{1\dots k-1} \otimes \pi)(dx_1, \dots, dx_k) &= \alpha_{1\dots k-1}(dx_1, \dots, dx_{k-1})\pi(x_{k-1}, dx_k) \\ &= \alpha(dx_1, \dots, dx_{k-1}, \mathbb{R}^d)\pi(x_{k-1}, dx_k). \end{aligned}$$

The proof of Theorem 3.1 is split into three parts: we prove the lower semi-continuity of the rate function (Proposition 4.1) then the lower bound (Proposition 5.1) and finally the upper bound (Proposition 6.1). We shall use the well-known fact that the process  $(X_{i-1}, X_i)_{i \in \mathbb{N}}$  is itself a Markov chain on  $\mathbb{R}^d \times \mathbb{R}^d$  with transition kernel

$$(3.9) \quad \pi^{(2)}(x_1, x_2, dy_1, dy_2) := \delta_{x_2}(y_1)\pi(x_2, dy_2), \quad x_1, x_2, y_1, y_2 \in \mathbb{R}^d.$$

#### 4. LOWER SEMI-CONTINUITY OF THE RATE FUNCTION

**Proposition 4.1.** *The function*

$$(4.1) \quad \xi \in \tilde{\mathcal{X}}^{(2)} \mapsto \tilde{J}^{(2)}(\xi) = \sum_{\alpha \in \xi} h(\alpha | \alpha_1 \otimes \pi)$$

*is lower semi-continuous.*

We recall that with a slight abuse of notation,  $\alpha_1$  denotes the projection of  $\alpha$  onto the first  $d$  coordinates in the proposition above.

*Proof of Proposition 4.1.* We use the notation  $J^{(2)}(\alpha) := h(\alpha | \alpha_1 \otimes \pi)$ , where  $\alpha \in \mathcal{M}_{\leq 1}^{(2)}$ . We will also use that

$$(4.2) \quad J^{(2)}(\alpha) = \sup_v \int \log \left( \frac{v}{\pi^{(2)}v} \right) d\alpha,$$

where the supremum runs over all bounded and Borel-measurable (or continuous and compactly supported) functions  $v: \mathbb{R}^d \times \mathbb{R}^d \rightarrow [1, +\infty)$ , see [6, Theorem 6.5.12 and Corollary 6.5.10]. Let  $(\mu_n)$  be a sequence in  $\tilde{\mathcal{X}}^{(2)}$  converging to  $\xi = \{\tilde{\alpha}_i\}_{i \in I}$ . Suppose

that there exists  $\ell \in (0, \infty)$  such that  $\tilde{J}^{(2)}(\mu_n) \leq \ell$  for all  $n$  large enough and let us show that  $\tilde{J}^{(2)}(\xi) \leq \ell$ . We for now restrict to the case where for each  $n$ ,  $\mu_n$  is made of a single orbit, so that  $\tilde{J}^{(2)}(\mu_n) = J^{(2)}(\mu_n)$ . Let  $\varepsilon > 0$ . By the same arguments used in Proposition 2.10 and possibly restricting to a subsequence, we may write for some  $k \geq 1$ ,

$$(4.3) \quad \mu_n = \sum_{i=1}^k \alpha_n^{(i)} + \beta_n,$$

where  $\alpha_n^{(i)}$  ( $1 \leq i \leq k$ ) and  $\beta_n$  are sequences of sub-probability measures in  $\mathbb{R}^d \times \mathbb{R}^d$  such that

$$(4.4) \quad \alpha_n^{(i)} * \delta_{(a_n^{(i)}, a_n^{(i)})} \Rightarrow \alpha_i \in \tilde{\alpha}_i, \quad n \rightarrow \infty,$$

for some sequences  $a_n^{(i)}$  ( $1 \leq i \leq k$ ) in  $\mathbb{R}^d$  satisfying

$$(4.5) \quad \lim_{n \rightarrow \infty} \min_{i \neq j} |a_n^{(i)} - a_n^{(j)}| = +\infty,$$

and  $(\beta_n)$  is widely separated from each  $(\alpha_n^{(i)})$ , with (recall (2.39))

$$(4.6) \quad \limsup_{n \rightarrow \infty} q_{\beta_n}(r) \leq \varepsilon, \quad \forall r > 0.$$

Moreover, by Lemma 2.3 and the construction in Step 2 of Proposition 2.10, there exists for each  $i$  a sequence  $(R_n^{(i)})_{n \in \mathbb{N}_0}$  tending to infinity such that

$$(4.7) \quad \begin{aligned} \text{Supp}(\alpha_n^{(i)}) &\subseteq B((-a_n^{(i)}, -a_n^{(i)}), R_n^{(i)}), \\ i \neq j &\implies B((-a_n^{(i)}, -a_n^{(i)}), R_n^{(i)}) \cap B((-a_n^{(j)}, -a_n^{(j)}), R_n^{(j)}) = \emptyset, \end{aligned}$$

and the support of  $\beta_n$  is contained in the complement of  $\bigcup_i B((-a_n^{(i)}, -a_n^{(i)}), R_n^{(i)})$ . In particular, the supports of  $\alpha_n^{(1)}, \alpha_n^{(2)}, \dots, \beta_n$  are all disjoint. We now define

$$(4.8) \quad u_i^{(n)}(x, y) = u_i(x + a_n^{(i)}, y + a_n^{(i)}), \quad x, y \in \mathbb{R}^d,$$

where, for each  $i$ ,  $u_i$  is a given non-negative continuous function with compact support. Consequently, the support of each  $u_i^{(n)}$  is contained in a compact ball in  $\mathbb{R}^{2d}$  centered around  $(-a_n^{(i)}, -a_n^{(i)})$ . In particular, recalling (3.9) there is  $R > 0$  such that for all  $1 \leq i \leq k$ ,

$$(4.9) \quad \begin{aligned} \text{Supp } u_i^{(n)} &\subseteq B((-a_n^{(i)}, -a_n^{(i)}), R), \\ \text{Supp } \pi^{(2)} u_i^{(n)} &\subseteq \mathbb{R}^d \times B(-a_n^{(i)}, R), \end{aligned}$$

hence

$$(4.10) \quad \left( \frac{1 + u_i^{(n)}}{1 + \pi^{(2)} u_i^{(n)}} \right)(x, y) \neq 1 \implies y \in B(-a_n^{(i)}, R).$$

We may now check that

$$(4.11) \quad \lim_{n \rightarrow \infty} \int \log \left( \frac{1 + u_i^{(n)}}{1 + \pi^{(2)} u_i^{(n)}} \right) d\alpha_n^{(j)} = \begin{cases} \int \log \left( \frac{1 + u_i}{1 + \pi^{(2)} u_i} \right) d\alpha_i & (i = j) \\ 0 & (i \neq j). \end{cases}$$

The case  $i = j$  follows from (4.4) and (4.8) together with the fact that  $\pi$  only depends on the difference of its arguments. Assume now that  $i \neq j$ . Note that, from what precedes, the integral on the left-hand side may be restricted to couples  $(x, y)$  such that  $y \in B(-a_n^{(i)}, R) \cap B(-a_n^{(j)}, R_n^{(j)})$ . If this set is non-empty, then  $|a_n^{(i)} - a_n^{(j)}| \leq R + R_n^{(j)}$ . This is not possible when  $n$  is large enough, since  $|a_n^{(i)} - a_n^{(j)}| > R_n^{(i)} + R_n^{(j)}$ , so the integral is eventually zero. Finally, letting

$$(4.12) \quad u^{(n)} = \sum_{1 \leq i \leq k} u_i^{(n)},$$

we obtain:

$$(4.13) \quad \begin{aligned} \ell &\geq \liminf_{n \rightarrow \infty} J^{(2)}(\mu_n) \geq \liminf_{n \rightarrow \infty} \int \log \left( \frac{1 + u^{(n)}}{1 + \pi^{(2)} u^{(n)}} \right) d\mu_n \\ &\geq \liminf_{n \rightarrow \infty} \sum_{1 \leq i, j \leq k} \int \log \left( \frac{1 + u_i^{(n)}}{1 + \pi^{(2)} u_i^{(n)}} \right) d\alpha_n^{(j)} \\ &\geq \sum_{1 \leq i \leq k} J^{(2)}(\alpha_i) - \varepsilon. \end{aligned}$$

Here, the contribution coming from  $\beta_n$  is zero because for sufficiently large  $n$  we have that  $\beta_n(dx, dy) = 0$  for  $y \in B(-a_n^{(i)}, R)$ , which together with (4.10) shows that the contribution is indeed zero. It now remains to send  $\varepsilon$  to zero and  $k$  to infinity. Finally, to treat the general case, that is when  $\mu_n$  has possibly more than one orbit, the idea is the same as in the last paragraph of [16, proof of Lemma 4.2].  $\blacksquare$

## 5. LOWER BOUND

**Proposition 5.1.** *For any open set  $G$  in  $\tilde{\mathcal{X}}^{(2)}$ ,*

$$(5.1) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n^{(2)} \in G) \geq - \inf_{\xi \in G} \tilde{J}^{(2)}(\xi).$$

*Proof of Proposition 5.1.* Let  $\xi = \{\alpha_i, i \in I\}$  be an element of  $\tilde{\mathcal{X}}^{(2)}$  such that  $\tilde{J}^{(2)}(\xi) < +\infty$ , hence  $h(\alpha_i | \alpha_{i,1} \otimes \pi) < +\infty$  for all  $i \in I$ . Let  $U$  be any open neighborhood of  $\xi$ . It is enough to prove that

$$(5.2) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n^{(2)} \in U) \geq -\tilde{J}^{(2)}(\xi).$$

We proceed as in [16, Lemma 4.3] and use the density of  $\tilde{\mathcal{M}}_1^{(2)}$  in  $\tilde{\mathcal{X}}^{(2)}$ , see Proposition 2.10. Let  $k \geq 1$  and consider the sequence  $(\mu_M)$  defined as in (2.37), except that we replace the totally disintegrating sequence  $(\lambda_M)$  by  $(\lambda_{M,1} \otimes \pi)$ , that is totally disintegrating as well, since for every  $x \in \mathbb{R}^d$  and  $r > 0$ ,  $(\lambda_{M,1} \otimes \pi)(B((x, x), r)) \leq \lambda_{M,1}(B(x, r))$  (recall that we use the product norm on  $\mathbb{R}^d \times \mathbb{R}^d$ ). Moreover, it is such that  $\tilde{\mu}_M$  still converges in  $\tilde{\mathcal{X}}^{(2)}$  to  $\xi$ . Note that  $\nu \in \mathcal{M}_{\leq 1}^{(2)} \mapsto h(\nu | \nu_1 \otimes \pi)$  is sub-additive as the supremum of linear functions, see (4.2), and that  $h(\lambda_{M,1} \otimes \pi | \lambda_{M,1} \otimes \pi) = 0$ . Therefore, we obtain

$$(5.3) \quad J^{(2)}(\tilde{\mu}_M) = h(\mu_M | \mu_{M,1} \otimes \pi) \leq \sum_{1 \leq i \leq k} h(\alpha_i | \alpha_{i,1} \otimes \pi) \leq J^{(2)}(\xi).$$

Thus, we have shown that there exists a sequence  $(\mu_M)_{M \in \mathbb{N}_0}$  in  $\widetilde{\mathcal{M}}_1^{(2)}$  which converges in  $\widetilde{\mathcal{X}}^{(2)}$  to  $\xi$  and is such that

$$(5.4) \quad \limsup_{M \rightarrow \infty} J^{(2)}(\widetilde{\mu}_M) \leq J^{(2)}(\xi).$$

The lower bound now follows from the standard Large Deviation lower bound of the pair empirical measure on  $\mathcal{M}_1^{(2)}$  (see [10] and the discussion in Section 1).  $\blacksquare$

## 6. UPPER BOUND

In this section we prove the following

**Proposition 6.1.** *For any closed set  $F$  in  $\widetilde{\mathcal{X}}^{(2)}$ ,*

$$(6.1) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\widetilde{L}_n^{(2)} \in F) \leq - \inf_{\xi \in F} \widetilde{J}^{(2)}(\xi).$$

Let  $\mathcal{U}$  denote the space of non-negative, continuous and compactly supported functions defined on  $\mathbb{R}^d \times \mathbb{R}^d$ . For any  $k \geq 1$ ,  $c > 0$ ,  $u_1, \dots, u_k \in \mathcal{U}$  and  $a_1, \dots, a_k \in \mathbb{R}^d$ , let  $g = g(u, c, R): \mathbb{R}^d \times \mathbb{R}^d \rightarrow (0, \infty)$  be defined by

$$(6.2) \quad g(x, y) = c + \sum_{i=1}^k u_i(x + a_i, y + a_i) \varphi\left(\frac{x + a_i}{R}, \frac{y + a_i}{R}\right), \quad x, y \in \mathbb{R}^d,$$

where  $\varphi$  is a smooth non-negative function such that  $0 \leq \varphi \leq 1$ ,  $\varphi = 1$  inside the unit ball and  $\varphi = 0$  outside the ball of radius two. For all  $\mu \in \mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$ , define

$$(6.3) \quad F(u, c, R, \mu) = \sup_{\substack{a_1, \dots, a_k \\ \min_{i \neq j} |a_i - a_j| \geq 4R}} \int_{\mathbb{R}^d \times \mathbb{R}^d} - \log \left( \frac{\pi^{(2)} g(x, y)}{g(x, y)} \right) \mu(dx, dy).$$

Since  $F(u, c, R, \cdot)$  is invariant under shifts of the form  $\mu \rightarrow \mu * \delta_{(x, x)}$ , we may lift it up to a function  $\widetilde{F}$  defined on  $\widetilde{\mathcal{M}}_1^{(2)}$ . In the sequel, we write

$$(6.4) \quad u_{i,R}(x, y) := u_i(x, y) \varphi(x/R, y/R), \quad x, y \in \mathbb{R}^d.$$

The proof of the upper bound follows from the following three lemmas:

**Lemma 6.2** (Sub-exponential growth). *For any choice of  $(u, c, R)$ , we have*

$$(6.5) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \exp(n \widetilde{F}(u, c, R, \widetilde{L}_n^{(2)})) \leq 0.$$

**Lemma 6.3** (Lower-semicontinuous extension). *If the sequence  $(\widetilde{\mu}_n)$  converges to  $\xi = (\widetilde{\alpha}_i)_{i \in I}$  in  $(\widetilde{\mathcal{X}}^{(2)}, \mathbf{D}_2)$ , then*

$$(6.6) \quad \liminf_{n \rightarrow \infty} \widetilde{F}(u, c, R, \widetilde{\mu}_n) \geq \widetilde{\Lambda}(u, c, R, \xi),$$

where

$$(6.7) \quad \widetilde{\Lambda}(u, c, R, \xi) := \sup_{\{\widetilde{\alpha}_1, \dots, \widetilde{\alpha}_k\} \subseteq \xi} \sum_{i=1}^k \sup_{b \in \mathbb{R}^d} \int - \log \left\{ \frac{\pi^{(2)}(c + u_{i,R})(x, y)}{(c + u_{i,R})(x, y)} \right\} \alpha_i(dx + b, dy + b).$$



**Remark 6.4.** Lemma 6.3 is analogous to Lemma 4.6 in [16]. However, the two suprema in (6.7) are not in the original paper. First, we add the supremum over  $b \in \mathbb{R}^d$  so that the quantity inside is a function of the orbit  $\tilde{\alpha}_i$  only rather than a function of a particular member of its orbit. This has however no consequence on the sequel of the argument in [16], since they later consider a supremum over functions (Lemma 4.7) allowing for arbitrary shifts. The other supremum is here to stress that an element of  $\xi$  is a collection of sub-probability orbits rather than a sequence.

**Lemma 6.5.** We have

$$(6.8) \quad J^{(2)}(\xi) = \sup_{\substack{c, R > 0, k \geq 1 \\ u_1, \dots, u_k \in \mathcal{U}}} \tilde{\Lambda}(u, c, R, \xi).$$

*Proof of Lemma 6.2.* Use that

$$(6.9) \quad -n \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \left( \frac{\pi^{(2)} g(x, y)}{g(x, y)} \right) L_n^{(2)}(dx, dy) = \log \prod_{i=1}^n \frac{g(X_{i-1}, X_i)}{\pi^{(2)} g(X_{i-1}, X_i)},$$

so that

$$(6.10) \quad \mathbb{E} \left[ \exp \left( n \int_{\mathbb{R}^d \times \mathbb{R}^d} -\log \left( \frac{\pi^{(2)} g(x, y)}{g(x, y)} \right) L_n^{(2)}(dx, dy) \right) \right] = \mathbb{E} \left[ \prod_{i=1}^n \frac{g(X_{i-1}, X_i)}{\pi^{(2)} g(X_{i-1}, X_i)} \right].$$

We write the product as

$$(6.11) \quad \prod_{i=1}^n \frac{g(X_{i-1}, X_i)}{\pi^{(2)} g(X_{i-1}, X_i)} = \frac{g(X_0, X_1)}{\pi^{(2)} g(X_{n-1}, X_n)} \prod_{i=1}^{n-1} \frac{g(X_i, X_{i+1})}{\pi^{(2)} g(X_{i-1}, X_i)},$$

and since  $g$  is bounded from below by  $c$  we see that

$$(6.12) \quad \mathbb{E} \left[ \prod_{i=1}^n \frac{g(X_{i-1}, X_i)}{\pi^{(2)} g(X_{i-1}, X_i)} \right] \leq \frac{\sup g}{c} \mathbb{E} \left[ \prod_{i=1}^{n-1} \frac{g(X_i, X_{i+1})}{\pi^{(2)} g(X_{i-1}, X_i)} \right],$$

and, by the fact that  $Y_i = (X_{i-1}, X_i)$  is a Markov chain and an induction argument the last expectation is one. This shows that the exponential rate of the right-hand side of (6.10) is zero. It therefore only remains to deal with the case in which in (6.10) an additional supremum is taken over  $a_1, \dots, a_k$  as in the statement of the result. This follows via a coarse graining argument. The idea is that by (3.1) it is exponentially unlikely that  $X$  travels in the time interval  $[0, n]$  to a distance  $\rho_n$ , which allows one to restrict the supremum over  $a_1, \dots, a_k$  to balls of radius  $\rho_n$ . In a very similar way as in [16, proof of Lemma 4.5] one may then conclude, so we omit the details. ■

*Proof of Lemma 6.3.* As it can be seen from the second step of the proof of Proposition 2.10, convergence in  $\mathbf{D}_2$  implies the existence of a decomposition

$$(6.13) \quad \mu_n = \sum_{j=1}^k \alpha_{n,j} * \delta_{(a_{n,j}, a_{n,j})} + \beta_n,$$

along subsequences as in (2.43), where, for all  $1 \leq j \leq k$

- $(a_{n,j})_{n \geq 1}$  is a sequence in  $\mathbb{R}^d$  satisfying

$$(6.14) \quad |a_{n,i} - a_{n,j}| \geq 4R$$

if  $n$  large enough and  $i \neq j$ ;

- $\alpha_{n,j} * \delta_{(a_{n,j}, a_{n,j})}$  converges weakly to  $\alpha_j$  as  $n \rightarrow \infty$ , where  $\alpha_j$  is some element in the orbit of  $\tilde{\alpha}_j$ ;
- $(\alpha_{n,j})$  and  $(\beta_n)$  are widely separated.

Recall (6.2). Choosing  $a_i = -a_{n,i}$  in the definition of  $g$ , we obtain

$$(6.15) \quad g(x, y) = c + \sum_{i=1}^k u_{i,R}(x - a_{n,i}, y - a_{n,i}).$$

By (6.14) and our assumption on  $\varphi$ , at most one term in the sum above can be nonzero. Also,

$$(6.16) \quad \pi^{(2)} u_{i,R}(x - a_{n,i}, y - a_{n,i}) = \int u_{i,R}(y - a_{n,i}, z - a_{n,i}) p(z - y) \lambda(dz)$$

is nonzero for at most one value of  $1 \leq i \leq k$  which is the same as in the above. We finally obtain:

$$(6.17) \quad \log \left( \frac{\pi^{(2)} g(x, y)}{g(x, y)} \right) = \sum_{i=1}^k \log \left( \frac{c + \pi^{(2)} u_{i,R}(x - a_{n,i}, y - a_{n,i})}{c + u_{i,R}(x - a_{n,i}, y - a_{n,i})} \right)$$

and we can conclude almost as in [16, Lemma 4.6]. As we already pointed out in Remark 6.4, there is more flexibility in choosing  $a_i$  (see (6.2)), which explains the additional supremum (over  $b$ ) in our statement, compared to [16, Lemma 4.6]. Indeed, let  $b_j \in \mathbb{R}^d$  for all  $1 \leq j \leq k$ . Then we may choose  $a_i = -a_{n,i} + b_i$  instead of  $a_i = -a_{n,i}$ . If we require that  $|a_{n,i} - a_{n,j}| \geq 4R + \max_{1 \leq j \leq k} |b_j|$  instead of simply  $|a_{n,i} - a_{n,j}| \geq 4R$ , then we finally get our claim. To see why the second supremum appears, note that if the number of elements in  $\xi$  is infinite then in (6.13) we can actually choose the  $k$  in the sum independent of the  $k$  in (6.2). Therefore adjusting the choice of  $a_1, \dots, a_k$  we can obtain any collection  $\{\tilde{\alpha}_1, \dots, \tilde{\alpha}_k\}$  in the limit. We therefore can conclude. ■

*Proof of Lemma 6.5.* As in [16, Lemma 4.7], we get

$$(6.18) \quad \sup_{\substack{c, R > 0, \\ u_1, \dots, u_k \in \mathcal{U}}} \tilde{\Lambda}(u, c, R, \xi) = \sum_{i=1}^k \sup_{\substack{c > 0, \\ u \in \mathcal{U}}} \int -\log \frac{\pi^{(2)}(c + u)(x, y)}{(c + u)(x, y)} \alpha_i(dx, dy).$$

The supremum in the sum actually coincides with  $J^{(2)}(\alpha_i)$ , see [6, Theorem 6.5.12 and Corollary 6.5.10]. ■

*Proof of Proposition 6.1.* The proof follows from Lemmas 6.2-6.5 in the exact same way as Proposition 4.4 in [16] follows from Lemmas 4.5-4.7 therein. ■

## 7. ADAPTATION TO RESCALED RANDOM WALKS

A small adaptation of the proof of Proposition 6.1 yields the same result also for a rescaled random walk. To precisely formulate the result we need to introduce more notation. Let  $(X_i)_{i \in \mathbb{N}_0}$  be a random walk in  $\mathbb{Z}^d$ . Assume that its step distribution is centered and square-integrable, with  $(1/d)\text{Id}$  as covariance matrix. Let  $(a_n)$  be a sequence of positive real numbers converging to  $+\infty$  and such that  $a_n^2 = o(n)$ . Let  $\varepsilon > 0$ . Define  $\ell := \ell(\varepsilon, n) = \lfloor \varepsilon a_n^2 \rfloor$  and

$$(7.1) \quad X_i^{\varepsilon, n} := \frac{X_{i\ell}}{a_n}, \quad i \in \mathbb{N}_0.$$

Denote by  $L_{n, \varepsilon}^{(2)}$  the corresponding pair empirical measure, that is

$$(7.2) \quad L_{n, \varepsilon}^{(2)} = \frac{1}{\lfloor n/\ell \rfloor} \sum_{i=1}^{\lfloor n/\ell \rfloor} \delta_{(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})}.$$

**Remark 7.1.** *The relevant scale for potential applications to the Swiss cheese model [17, 18] corresponds to the choice  $a_n := n^{1/d}$ .*

We prove the following:

**Theorem 7.2.** *For any closed set  $F$  in  $\tilde{\mathcal{X}}^{(2)}$  and  $\varepsilon > 0$ ,*

$$(7.3) \quad \limsup_{n \rightarrow \infty} \frac{a_n^2}{n} \log \mathbb{P}(\tilde{L}_{n, \varepsilon}^{(2)} \in F) \leq -\frac{1}{\varepsilon} \inf_{\xi \in F} \tilde{J}_{\varepsilon/d}^{(2)}(\xi).$$

where  $\tilde{J}_{\varepsilon}^{(2)}$  is defined as in (3.5) with  $\pi$  replaced by  $\pi_{\varepsilon}$ , that is the Brownian semi-group at time  $\varepsilon$ .

*Proof of Theorem 7.2.* It turns out that only the proof and the statement of Lemma 6.2 need to be adapted. The remaining statements are about the rate function rather than the Markov chain at hand. To that end we define

$$(7.4) \quad \pi_{n, \varepsilon}^{(2)} g(x, y) = \mathbb{E} \left[ g(y, y + X_1^{\varepsilon, n}) \right], \quad \text{and} \quad \pi_{\varepsilon}^{(2)} g(x, y) = \mathbb{E} \left[ g(y, y + B_{\varepsilon}) \right].$$

With these notations at hand we define  $F$  as in Section 6 but with  $\pi^{(2)}$  replaced by  $\pi_{\varepsilon}^{(2)}$ . Then, defining  $M := \frac{n}{\ell} = \frac{n}{\lfloor \varepsilon a_n^2 \rfloor}$  we show that

$$(7.5) \quad \limsup_{n \rightarrow \infty} \frac{1}{M} \log \mathbb{E} \exp(M \tilde{F}(u, c, R, \tilde{L}_{n, \varepsilon}^{(2)})) \leq 0.$$

Note that the result would be immediate from the proof if  $\pi_{\varepsilon}^{(2)}$  would be replaced by the transition kernel  $\pi_{n, \varepsilon}^{(2)}$  of  $X^{\varepsilon, n}$ . Following the proof of Lemma 6.2, we write

$$(7.6) \quad \prod_{i=1}^M \frac{g(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})}{\pi_{\varepsilon}^{(2)} g(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})} = \prod_{i=1}^M \frac{g(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})}{\pi_{n, \varepsilon}^{(2)} g(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})} \prod_{i=1}^M \frac{\pi_{n, \varepsilon}^{(2)} g(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})}{\pi_{\varepsilon}^{(2)} g(X_{i-1}^{\varepsilon, n}, X_i^{\varepsilon, n})}.$$

Since (i)  $g$  is positive, continuous and constant outside of a compact set and (ii)  $X_1^{\varepsilon,n}$  converges in law to  $B_{\varepsilon/d}$  as  $n \rightarrow \infty$ , it follows that

$$(7.7) \quad \lim_{n \rightarrow \infty} \sup_{x,y \in \mathbb{R}^d} \left| \frac{\pi_{n,\varepsilon}^{(2)} g(x,y)}{\pi_{\varepsilon/d}^{(2)} g(x,y)} - 1 \right| = 0.$$

The variable  $x$  in the supremum plays no actual role and uniformity in  $y$  can be deduced from the uniform continuity of  $g$  and a coupling for which  $X_1^{\varepsilon,n}$  converges to  $B_\varepsilon$  almost surely as  $n \rightarrow \infty$ . The convergence in (7.7) allows to control the rightmost factor in (7.6). The first factor on the right-hand side can be dealt with as in the proof of Lemma 6.2. ■

## 8. FROM THE PAIR EMPIRICAL MEASURE TO THE EMPIRICAL MEASURE

Consider a Markov chain  $(X_i)_{i \in \mathbb{N}_0}$  satisfying the same assumptions as in Section 3 and define the empirical measure  $L_n$  of  $(X_i)_{i \in \mathbb{N}_0}$  by

$$(8.1) \quad L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

Note that  $L_n$  is simply the second marginal of  $L_n^{(2)}$ . We further denote by  $\tilde{L}_n$  the orbit of  $L_n$  in  $\tilde{\mathcal{M}}_1(\mathbb{R}^d)$ , the space defined in [16, Section 2], see Remark 2.1 for the difference between  $\tilde{\mathcal{M}}_1^{(2)}$  and  $\tilde{\mathcal{M}}_1(\mathbb{R}^{2d})$ . We moreover denote by  $\tilde{\mathcal{X}}$  the compactification of  $\tilde{\mathcal{M}}_1(\mathbb{R}^d)$ , equipped with the metric  $\mathbf{D}$ , see [16, Section 3] for details or Appendix A for a short overview. Given a sub-probability measure  $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R}^d \times \mathbb{R}^d)$  we denote by  $\text{Pr}(\mu)$  its projection onto the second marginal (of course, the sequel applies to the projection onto the first marginal as well). We then define a projection map  $\tilde{\text{Pr}} : \tilde{\mathcal{X}}^{(2)} \rightarrow \tilde{\mathcal{X}}$  by first defining

$$(8.2) \quad \tilde{\text{Pr}}(\tilde{\mu}) = \{\text{Pr}(\mu) * \delta_x : x \in \mathbb{R}^d\}$$

for  $\tilde{\mu} \in \tilde{\mathcal{M}}_{\leq 1}(\mathbb{R}^d \times \mathbb{R}^d)$ , then

$$(8.3) \quad \tilde{\text{Pr}}(\xi) = \{\tilde{\text{Pr}}(\tilde{\alpha}_i)\}_{i \in I}$$

for  $\xi = \{\tilde{\alpha}_i\}_{i \in I} \in \tilde{\mathcal{X}}^{(2)}$ . The goal of this section is to derive a Large Deviation principle for  $\tilde{L}_n$  from one of  $\tilde{L}_n^{(2)}$ , which is done in Theorem 8.5 below. By the contraction principle it would be sufficient to check continuity of the projection map. This unfortunately turns out to be false as the following counter-example shows. Consider the sequence of measures given by  $\mu_n = \delta_{(n,0)}$ . Then, on the one hand we have that for any  $f \in \mathcal{F}_k^{(2)}$

$$(8.4) \quad \int f(u_1, v_1, \dots, u_k, v_k) \prod_{1 \leq i \leq k} \delta_{(n,0)}(du_i, dv_i) = f(n, 0, \dots, n, 0).$$

Since  $f \in \mathcal{F}_k^{(2)}$  the latter quantity converges to zero as  $n \rightarrow \infty$ . This shows that the sequence  $\tilde{\mu}_n$  converges to  $\emptyset$  in  $(\tilde{\mathcal{X}}^{(2)}, \mathbf{D}_2)$ . On the other hand,

$$(8.5) \quad \tilde{\text{Pr}}(\tilde{\mu}_n) = \{\delta_x : x \in \mathbb{R}^d\},$$

which is a constant sequence and therefore converges to  $\{\delta_x : x \in \mathbb{R}^d\}$ . In particular,

$$(8.6) \quad \lim_{n \rightarrow \infty} \widetilde{\Pr}(\widetilde{\mu}_n) \neq \widetilde{\Pr}(\lim_{n \rightarrow \infty} \widetilde{\mu}_n),$$

which shows that  $\widetilde{\Pr}$  is not continuous and we cannot apply the standard contraction principle. There are of course generalizations of the classical contraction principle, however it is not clear to us how to use them in the present context. The key to still obtain a Large Deviation principle is via a smoothing of the projection map. More precisely, for  $A > 0$ , we define  $\phi_A : \mathbb{R}^d \rightarrow [0, 1]$  to be a smooth function such that

$$(8.7) \quad \phi_A(x) = \begin{cases} 1, & |x| \leq A, \\ 0, & |x| \geq 2A. \end{cases}$$

We then define for any sub-probability measure  $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R}^d \times \mathbb{R}^d)$

$$(8.8) \quad (\Pr_A \mu)(B) = \int_{v \in B} \int_{u \in \mathbb{R}^d} \phi_A(u - v) \mu(du, dv).$$

As the reader may check,

$$(8.9) \quad \Pr_A(\mu * \delta_{(x,x)}) = (\Pr_A \mu) * \delta_x, \quad x \in \mathbb{R}^d,$$

so  $\Pr_A$  can be naturally extended to a map  $\widetilde{\Pr}_A : \widetilde{\mathcal{X}}^{(2)} \mapsto \widetilde{\mathcal{X}}$ . We then first observe that:

**Proposition 8.1.** *The map  $\widetilde{\Pr}_A$  is continuous from  $(\widetilde{\mathcal{X}}^{(2)}, \mathbf{D}^{(2)})$  to  $(\widetilde{\mathcal{X}}, \mathbf{D})$ .*

*Proof of Proposition 8.1.* Given  $\xi \in \widetilde{\mathcal{X}}^{(2)}$  and  $f \in \mathcal{F}_k$  (in particular  $f$  is bounded), we have that

$$(8.10) \quad \sum_{\alpha \in \xi} \int f(v_1, \dots, v_k) \prod_{i=1}^k (\Pr_A \alpha)(dv_i) = \sum_{\alpha \in \xi} \int f(v_1, \dots, v_k) \prod_{i=1}^k \phi_A(u_i - v_i) \alpha(du_i, dv_i).$$

To conclude it only remains to observe that

$$(8.11) \quad V_A : (u_1, v_1, \dots, u_k, v_k) \mapsto f(v_1, \dots, v_k) \prod_{i=1}^k \phi_A(u_i - v_i) \in \mathcal{F}_k^{(2)}.$$

Indeed,  $V_A$  is clearly continuous and it is also vanishing: if we write, for convenience,

$$(8.12) \quad (u_1, v_1, \dots, u_k, v_k) = (x_1, \dots, x_{2k}),$$

and assume that  $\max\{|x_i - x_j| : i \neq j\} \geq M$ , then one of the four following cases indeed occurs:

- (1)  $|u_i - v_i| \geq M$  for some index  $i$ ;
- (2)  $|v_i - v_j| \geq M$  for some index pair  $i \neq j$ ;
- (3)  $|u_i - v_j| \geq M$  for some index pair  $i \neq j$ ;
- (4)  $|u_i - u_j| \geq M$  for some index pair  $i \neq j$ .

In the first case, we write

$$(8.13) \quad V_A(u_1, v_1, \dots, u_k, v_k) \leq \|f\|_\infty \sup\{\phi_A(x) : |x| \geq M\},$$

which is eventually zero when  $M$  is large enough, since  $\phi_A$  is compactly supported. In the second case, we write

$$(8.14) \quad V_A(u_1, v_1, \dots, u_k, v_k) \leq \sup\left\{f(v_1, \dots, v_k) : \max_{i \neq j} |v_i - v_j| \geq M\right\},$$

which decreases to zero as  $M \rightarrow \infty$ , since  $f$  itself is vanishing. In the third case, we use the triangular inequality  $|u_i - v_j| \leq |u_i - v_i| + |v_i - v_j|$  so that we are back to one of the first two cases with  $M/2$  instead of  $M$ . In the fourth case, we use the triangular inequality  $|u_i - u_j| \leq |u_i - v_j| + |v_j - u_j|$  so that we are back to either the first or third case with  $M/2$  instead of  $M$ . Hence, the proof is complete.  $\blacksquare$

To continue we further define the diagonal with width  $A$  in  $\mathbb{R}^d \times \mathbb{R}^d$  via

$$(8.15) \quad \mathcal{D}_A := \{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d : |u - v| \leq A\},$$

along with

$$(8.16) \quad \mathcal{M}(A, \delta) := \{\xi \in \tilde{\mathcal{X}}^{(2)} : \xi(\mathcal{D}_A) \geq 1 - \delta\},$$

where

$$(8.17) \quad \xi(\mathcal{D}_A) = \sum_{\alpha \in \xi} \alpha(\mathcal{D}_A).$$

The latter is well defined, i.e., it does not depend on the choice of the orbits. In Lemmas 8.2 and 8.3 below, we make two observations on the set  $\mathcal{M}(A, \delta)$  that shall prove useful in the sequel.

**Lemma 8.2.**  $\mathcal{M}(A, \delta)$  is a closed subset of  $\tilde{\mathcal{X}}^{(2)}$ .

*Proof of Lemma 8.2.* Let  $(\xi_m)$  be a sequence in  $\mathcal{M}(A, \delta)$  converging with respect to the metric  $\mathbf{D}_2$  to some element  $\xi$ . We need to show that  $\xi \in \mathcal{M}(A, \delta)$ . To that end we define a collection of smooth functions  $\phi_{A_1, A_2} : [0, \infty) \mapsto [0, 1]$ , where  $A < A_1 < A_2$ , with the following properties

$$(8.18) \quad \phi_{A_1, A_2}(x) = \begin{cases} 1, & A_1 < x < A_2, \\ 0, & x < \frac{1}{2}(A + A_1) \text{ or } x > 2A_2, \end{cases}$$

and such that  $\phi_{A_1, A_2}$  monotonically increases to the indicator of  $(A, \infty)$  (pointwise) as  $A_1$  and  $A_2$  tend to  $A$  and  $+\infty$  respectively. Given any sub-probability measure  $\alpha \in \mathcal{M}_{\leq 1}(\mathbb{R}^d \times \mathbb{R}^d)$  we then see that by the monotone convergence theorem

$$(8.19) \quad \alpha(\mathcal{D}_A^c) = \sup_{A < A_1 < A_2} \int \phi_{A_1, A_2}(|u - v|) \alpha(du, dv).$$

We now introduce another collection of smooth functions  $\varphi_B : [0, \infty) \rightarrow [0, 1]$  such that  $\varphi_B(x) = 1$  if  $x \leq B$  and  $\varphi_B(x) = 0$  if  $x > 2B$ , and such that  $\varphi_B$  monotonically increases

(pointwise) to the constant function equal to one as  $B$  tends to infinity. Again, by the monotone convergence theorem:

$$(8.20) \quad \alpha(\mathcal{D}_A^c) = \sup_{\substack{A < A_1 < A_2 \\ B > 0}} I(\alpha; A_1, A_2, B).$$

where

$$(8.21) \quad I(\alpha; A_1, A_2, B) = \int \phi_{A_1, A_2}(|u_1 - v_1|) \varphi_B(|u_1 - u_2|) \varphi_B(|v_1 - v_2|) \alpha(du_1, dv_1) \alpha(du_2, dv_2).$$

We finally obtain:

$$(8.22) \quad \begin{aligned} \sum_{\alpha \in \xi} \alpha(\mathcal{D}_A^c) &= \sup_{\substack{A < A_1 < A_2 \\ B > 0}} \sum_{\alpha \in \xi} I(\alpha; A_1, A_2, B) \\ &= \sup_{\substack{A < A_1 < A_2 \\ B > 0}} \lim_{m \rightarrow \infty} \sum_{\alpha \in \xi_m} I(\alpha; A_1, A_2, B) \\ &\leq \liminf_{m \rightarrow \infty} \sum_{\alpha \in \xi_m} \alpha(\mathcal{D}_A^c) \leq \delta. \end{aligned}$$

On the second line we used the convergence of  $\xi_m$  to  $\xi$  along with the fact that the function

$$(8.23) \quad (u_1, v_1, u_2, v_2) \in (\mathbb{R}^d)^4 \mapsto \phi_{A_1, A_2}(|u_1 - v_1|) \varphi_B(|u_1 - u_2|) \varphi_B(|v_1 - v_2|)$$

is smooth, translation invariant and vanishing, since  $\phi_{A_1, A_2}$  and  $\varphi_B$  are compactly supported. This completes the proof.  $\blacksquare$

**Lemma 8.3.** *For every  $A, \delta > 0$ ,*

$$(8.24) \quad \mathcal{M}(A, \delta) \subseteq \{\xi \in \tilde{\mathcal{X}}^{(2)} : \mathbf{D}(\tilde{\text{Pr}}(\xi), \tilde{\text{Pr}}_A(\xi)) \leq 2\delta\}.$$

*Proof of Lemma 8.3.* We first estimate the distance between  $\tilde{\text{Pr}}_A$  and  $\tilde{\text{Pr}}$ . Given any function  $f \in \mathcal{F}_k$  and any sub-probability measure  $\mu \in \mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$ ,

$$(8.25) \quad \begin{aligned} &\int f(v_1, \dots, v_k) \left[ \prod_{i=1}^k (\text{Pr}\mu)(dv_i) - \prod_{i=1}^k (\text{Pr}_A\mu)(dv_i) \right] \\ &= \sum_{j=1}^k \int f(v_1, \dots, v_k) [(\text{Pr}\mu - \text{Pr}_A\mu)(dv_j)] \prod_{i < j} (\text{Pr}\mu)(dv_i) \prod_{i > j} (\text{Pr}_A\mu)(dv_i) \\ &= \sum_{j=1}^k \int f(v_1, \dots, v_k) [(1 - \phi_A)(u_j - v_j) \mu(du_j, dv_j)] \prod_{i < j} (\text{Pr}\mu)(dv_i) \prod_{i > j} (\text{Pr}_A\mu)(dv_i). \end{aligned}$$

By (8.7), the above can be bounded from above (in absolute value) by

$$(8.26) \quad k \|f\|_\infty \times \mu(\{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d : |u - v| \geq A\}).$$

From the definition of the metric  $\mathbf{D}$  in (A.7) and [16, Section 3, Eq. (17)], we may deduce that

$$(8.27) \quad \mathbf{D}(\tilde{\text{Pr}}(\xi), \tilde{\text{Pr}}_A(\xi)) \leq 2 \times \xi(\mathcal{D}_A^c).$$

■

Our final key observation before we state the main result of this section is the following:

**Proposition 8.4.** *The sequence  $(\tilde{L}_n^{(2)})$  is diagonally exponentially tight, i.e. for all  $\delta > 0$*

$$(8.28) \quad \lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n^{(2)} \notin \mathcal{M}(A, \delta)) = -\infty.$$

*Proof of Proposition 8.4.* This follows from:

$$(8.29) \quad \mathbb{P}(\tilde{L}_n^{(2)} \notin \mathcal{M}(A, \delta)) = \mathbb{P}(\text{Bin}(n, \mathbb{P}(|X_1| \geq A)) > n\delta),$$

where we recall that  $X_1$  is distributed as a single step of the underlying Markov chain. ■

Under the above assumption we can now formulate the following result:

**Theorem 8.5.** *Under the assumptions (1) to (3) in Section 3, the sequence  $(\tilde{L}_n)$  satisfies a Large Deviation principle in  $\tilde{\mathcal{X}}$  with rate  $n$  and rate function*

$$(8.30) \quad \tilde{J}(\xi) = \sum_{\alpha \in \xi} J(\alpha),$$

where

$$(8.31) \quad J(\alpha) = \inf\{J^{(2)}(\mu) : \mathbb{P}r(\mu) = \alpha\}.$$

Before we proceed with the proof, let us remark that the rate function  $\tilde{J}$  is actually what we would get from a blunt application of the contraction principle.

**Lemma 8.6.** *Let  $\tilde{J}$  be defined as in (8.41) and  $\xi \in \tilde{\mathcal{X}}$ . Then,*

$$(8.32) \quad \tilde{J}(\xi) = \inf\{\tilde{J}^{(2)}(\theta) : \widetilde{\mathbb{P}r}(\theta) = \xi\}.$$

Moreover,  $\tilde{J}$  is a lower semi-continuous function.

*Proof of Lemma 8.6.* Write  $\xi = \{\alpha_i : i \in I\}$ . Note that for every  $\theta$  such that  $\widetilde{\mathbb{P}r}(\theta) = \xi$ ,

$$(8.33) \quad \sum_{i \in I} J(\alpha_i) \leq \tilde{J}^{(2)}(\theta).$$

Indeed, writing  $\theta = \{\nu_i : i \in I\}$  we have  $\widetilde{\mathbb{P}r}(\nu_i) = \alpha_i$  for every  $i \in I$ , hence  $\tilde{J}^{(2)}(\theta) = \sum_{i \in I} J^{(2)}(\nu_i) \geq \sum_{i \in I} J(\alpha_i)$ . Therefore,  $\tilde{J}(\xi) \geq \sum_{i \in I} J(\alpha_i)$ . We now proceed with the upper bound. We can assume that the index set  $I$  is a subset of the natural numbers or coincides with it. Fix  $\varepsilon > 0$ . For all  $i \in I$ , there exists  $\nu_i^{(\varepsilon)}$  such that  $\mathbb{P}r(\nu_i^{(\varepsilon)}) = \alpha_i$  and  $|J^{(2)}(\nu_i^{(\varepsilon)}) - J(\alpha_i)| \leq \varepsilon 2^{-i}$ . Define  $\theta^{(\varepsilon)} = \{\nu_i^{(\varepsilon)}, i \in I\}$ . Then  $\widetilde{\mathbb{P}r}\theta^{(\varepsilon)} = \xi$ , from which it follows that:

$$(8.34) \quad \tilde{J}(\xi) \leq \tilde{J}^{(2)}(\theta^{(\varepsilon)}) = \sum_{i \in I} J^{(2)}(\nu_i^{(\varepsilon)}) \leq \sum_{i \in I} J(\alpha_i) + \varepsilon.$$

To show that  $\tilde{J}$  is lower semi-continuous, we first remark that the function  $J$  appearing in (8.30) may be written as

$$(8.35) \quad J(\alpha) = \sup_v \int \log\left(\frac{v}{\pi v}\right) d\alpha,$$



where the supremum runs over all bounded and Borel-measurable (or continuous and compactly supported) functions  $v: \mathbb{R}^d \rightarrow [1, +\infty)$ . This standard fact may be inferred, for instance, from [9, Lemma 2.1 of the first paper] and [10, Theorem 2.1]. From this variational representation, one can mimick the proof of Proposition 4.1, that is itself inspired by [16, Lemma 4.2].  $\blacksquare$

*Proof of Theorem 8.5.* We start with the proof of the upper bound. Let  $F \subseteq \tilde{\mathcal{X}}$  be a closed set. We can then estimate

$$(8.36) \quad \begin{aligned} \mathbb{P}(\tilde{L}_n \in F) &\leq \mathbb{P}(\tilde{L}_n \in F, \tilde{L}_n^{(2)} \in \mathcal{M}(A, \delta)) + \mathbb{P}(\tilde{L}_n^{(2)} \notin \mathcal{M}(A, \delta)) \\ &\leq \mathbb{P}(\tilde{L}_n^{(2)} \in \widetilde{\text{Pr}}_A^{-1}(F_{2\delta}), \tilde{L}_n^{(2)} \in \mathcal{M}(A, \delta)) + \mathbb{P}(\tilde{L}_n^{(2)} \notin \mathcal{M}(A, \delta)). \end{aligned}$$

Here, for any  $a > 0$  we defined the closed set

$$(8.37) \quad F_a = \{\xi \in \tilde{\mathcal{X}} : \mathbf{D}(\xi, F) \leq a\},$$

and the second inequality above follows from Lemma 8.3. Since  $F_{2\delta}$  is closed and  $\widetilde{\text{Pr}}_A$  is continuous it follows that  $\widetilde{\text{Pr}}_A^{-1}(F_{2\delta})$  is closed. For every  $A, \delta > 0$ , we note:

$$(8.38) \quad -C(A, \delta) := \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n^{(2)} \notin \mathcal{M}(A, \delta)).$$

Using Lemma 8.2 and the Large Deviation principle of the pair empirical measure we can conclude that

$$(8.39) \quad \begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n \in F) &\leq -\left( C(A, \delta) \wedge \inf_{\xi \in \mathcal{M}(A, \delta) \cap \widetilde{\text{Pr}}_A^{-1}(F_{2\delta})} \tilde{J}^{(2)}(\xi) \right) \\ &\leq -\left( C(A, \delta) \wedge \inf_{\xi \in \mathcal{M}(A, \delta) \cap \widetilde{\text{Pr}}^{-1}(F_{4\delta})} \tilde{J}^{(2)}(\xi) \right). \end{aligned}$$

here, the last inequality is a consequence of Lemma 8.3. Note that by Proposition 8.4 the constant  $C(A, \delta)$  tends to infinity when  $A$  tends to infinity. Therefore, we see that for any  $\delta > 0$

$$(8.40) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n \in F) \leq - \inf_{\xi \in F_\delta} \tilde{J}(\xi),$$

where

$$(8.41) \quad \tilde{J}(\xi) := \inf\{\tilde{J}^{(2)}(\theta) : \widetilde{\text{Pr}}(\theta) = \xi\}.$$

The fact that

$$(8.42) \quad \tilde{J}(\xi) = \sum_{\alpha \in \xi} J(\alpha),$$

and that  $\tilde{J}$  is lower semi-continuous are consequences of Lemma 8.6 above. To finish the derivation of the Large Deviation upper bound we need to send  $\delta$  to 0. Since  $F_\delta$  is a closed set of a compact space it is compact itself. Since  $\tilde{J}$  is lower semi-continuous there exists  $\xi_\delta \in F_\delta$  minimizing  $\tilde{J}$  over  $F_\delta$ . By compactness of  $\tilde{\mathcal{X}}$  we can extract a subsequence along which  $\xi_\delta$  converges as  $\delta \rightarrow 0$ . For ease of notation we denote this subsequence by  $\xi_\delta$  and its limit by  $\xi_0$ . By the lower semi-continuity of  $\tilde{J}$  we have that

$$(8.43) \quad \tilde{J}(\xi_0) \leq \liminf_{\delta \rightarrow 0} \tilde{J}(\xi_\delta).$$

Since  $\xi \mapsto \mathbf{D}(\xi, F)$  is continuous it follows that  $\xi_0 \in F$ . Therefore,

$$(8.44) \quad \inf_{\xi \in F} \tilde{J}(\xi) \leq \tilde{J}(\xi_0) \leq \liminf_{\delta \rightarrow 0} \tilde{J}(\xi_\delta),$$

which allows to conclude the upper bound.

We now come to the proof of the lower bound. Let  $U$  be an open set in  $\tilde{\mathcal{X}}$  w.r.t the metric  $\mathbf{D}$ . We first show that

$$(8.45) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n \in U) \geq - \liminf_{A \rightarrow \infty} \inf_{\tilde{\text{Pr}}_A^{-1}(B(\xi, \delta))} \tilde{J}^{(2)},$$

where  $\xi \in U$  and  $\delta$  small enough such that  $B(\xi, \delta) \subseteq U$ . To see that we can write using Lemma 8.3,

$$(8.46) \quad \begin{aligned} \mathbb{P}(\tilde{L}_n \in U) &\geq \mathbb{P}(\tilde{L}_n \in B(\xi, \delta), \tilde{L}_n^{(2)} \in \mathcal{M}(A, \delta/4)) \\ &\geq \mathbb{P}(\tilde{\text{Pr}}_A(\tilde{L}_n^{(2)}) \in B(\xi, \delta/2), \tilde{L}_n^{(2)} \in \mathcal{M}(A, \delta/4)) \\ &\geq \mathbb{P}(\tilde{\text{Pr}}_A(\tilde{L}_n^{(2)}) \in B(\xi, \delta/2)) - \mathbb{P}(\tilde{L}_n^{(2)} \notin \mathcal{M}(A, \delta/4)). \end{aligned}$$

Applying the Large Deviation Principle for  $\tilde{L}_n^{(2)}$  and Proposition 8.4 allows to conclude (8.45). To continue note that by Equation (8.27) for any  $\xi^{(2)} \in \tilde{\mathcal{X}}^{(2)}$  one has that  $\mathbf{D}(\tilde{\text{Pr}}(\xi^{(2)}), \tilde{\text{Pr}}_A(\xi^{(2)})) \leq 2\xi^{(2)}(\mathcal{D}_A^c)$ , which tends to zero as  $A \rightarrow \infty$ . Indeed, note that

$$(8.47) \quad \xi^{(2)}(\mathcal{D}_A^c) = \sum_{\alpha \in \xi^{(2)}} \alpha(\mathcal{D}_A^c),$$

and each term in the sum tends to zero as  $A \rightarrow \infty$ . Since moreover,  $\alpha(\mathcal{D}_A^c) \leq \alpha(\mathbb{R}^d \times \mathbb{R}^d)$  and the latter is summable over all the elements in  $\xi$ , we can conclude using the dominated convergence theorem. Hence, given any  $\xi^{(2)} \in \tilde{\mathcal{X}}^{(2)}$  such that  $\tilde{\text{Pr}}(\xi^{(2)}) = \xi$  there exists  $A_0 > 0$  such that for all  $A \geq A_0$  one has that  $\tilde{\text{Pr}}_A(\xi^{(2)}) \in B(\xi, \delta)$ . We define

$$(8.48) \quad \mathcal{C}_A = \{\xi^{(2)} \in \tilde{\mathcal{X}}^{(2)} : \tilde{\text{Pr}}(\xi^{(2)}) = \xi, \tilde{\text{Pr}}_A(\xi^{(2)}) \in B(\xi, \delta)\}.$$

It follows directly from the definition of  $\mathcal{C}_A$  that

$$(8.49) \quad \inf_{\tilde{\text{Pr}}_A^{-1}(B(\xi, \delta))} \tilde{J}^{(2)} \leq \inf_{\mathcal{C}_A} \tilde{J}^{(2)}.$$

Our goal is to prove that

$$(8.50) \quad \limsup_{A \rightarrow \infty} \inf_{\mathcal{C}_A} \tilde{J}^{(2)} \leq \inf_{\text{Pr}^{-1}(\{\xi\})} \tilde{J}^{(2)},$$

which would yield the result. Let  $\varepsilon > 0$ , and fix  $\xi_\varepsilon^{(2)} \in \tilde{\mathcal{X}}^{(2)}$  such that  $\text{Pr}(\xi_\varepsilon^{(2)}) = \xi$  and such that additionally

$$(8.51) \quad \inf_{\text{Pr}^{-1}(\{\xi\})} \tilde{J}^{(2)} \geq \tilde{J}^{(2)}(\xi_\varepsilon^{(2)}) - \varepsilon.$$

By the above, there exists  $A_0 > 0$  such that for all  $A \geq A_0$  one has that  $\xi_\varepsilon^{(2)} \in \mathcal{C}_A$ . This allows us to conclude that

$$(8.52) \quad \inf_{\text{Pr}^{-1}(\{\xi\})} \tilde{J}^{(2)} \geq \tilde{J}^{(2)}(\xi_\varepsilon^{(2)}) - \varepsilon \geq \inf_{\mathcal{C}_A} \tilde{J}^{(2)} - \varepsilon$$

for all  $A \geq A_0$ . Sending  $A$  to infinity and  $\varepsilon$  to zero allows us to deduce (8.50). Hence, given  $\xi \in \tilde{\mathcal{X}}$  and any open set  $U$  containing  $\xi$  we can conclude that

$$(8.53) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\tilde{L}_n \in U) \geq - \inf_{\tilde{\text{Pr}}^{-1}(\{\xi\})} \tilde{J}^{(2)} = -\tilde{J}(\xi),$$

and the result follows. To finish the proof of Theorem 8.5 we need to establish the lower semi-continuity of the rate function. This however is the content of Lemma 8.6. Hence, we can conclude.  $\blacksquare$

**8.1. Additional observations.** Let us come back to the counter-example provided at the beginning of this section and formulate a few observations that may be of independent interest. Looking back at (8.6), we remark that the left-hand side therein is in some sense “larger” (i.e. contains more mass) than the right-hand side. We formalize this sort of “lower semi-continuity” with the following:

**Proposition 8.7.** *Assume that the sequence  $(\xi_n)$  in  $\tilde{\mathcal{X}}^{(2)}$  converges to  $\xi$  w.r.t to the metric  $\mathbf{D}^{(2)}$ . Then, for every non-negative function  $f \in \mathcal{F}_2$ ,*

$$(8.54) \quad \liminf_{n \rightarrow \infty} \int f(v_1, v_2) \tilde{\text{Pr}}\xi_n(dv_1) \tilde{\text{Pr}}\xi_n(dv_2) \geq \int f(v_1, v_2) \tilde{\text{Pr}}\xi(dv_1) \tilde{\text{Pr}}\xi(dv_2).$$

*Proof of Proposition 8.7.* We use the smooth truncation functions from (8.7). Then, for all  $A > 0$ ,

$$(8.55) \quad \begin{aligned} \int f(v_1, v_2) \tilde{\text{Pr}}\xi_n(dv_1) \tilde{\text{Pr}}\xi_n(dv_2) &= \int f(v_1, v_2) \xi_n(du_1, dv_1) \xi_n(du_2, dv_2) \\ &\geq \int f_A(u_1, v_1, u_2, v_2) \xi_n(du_1, v_1) \xi_n(du_2, dv_2), \end{aligned}$$

where

$$(8.56) \quad f_A(u_1, v_1, u_2, v_2) := f(v_1, v_2) \phi_A(u_1 - u_2).$$

Repeating the arguments between (8.11) and (8.14), one can show that  $f_A \in \mathcal{F}_2^{(2)}$ . Therefore,

$$(8.57) \quad \liminf_{n \rightarrow \infty} \int f(v_1, v_2) \tilde{\text{Pr}}\xi_n(dv_1) \tilde{\text{Pr}}\xi_n(dv_2) \geq \int f_A(u_1, v_1, u_2, v_2) \xi(du_1, v_1) \xi(du_2, dv_2)$$

Letting  $A \rightarrow \infty$  in the right-hand side above and using the monotone convergence theorem completes the proof.  $\blacksquare$

The proof above indicates that the concept of *diagonal tightness* introduced above should be sufficient to retrieve continuity of the projection map. This is the purpose of the below proposition, in which we use the following notation: if  $\xi = \{\tilde{\alpha}_i\}_{i \in I} \in \tilde{\mathcal{X}}^{(2)}$  and if  $A$  is a *diagonally shift invariant* Borel subset of  $\mathbb{R}^d$  (meaning that  $A + (x, x) = A$  for every  $x \in \mathbb{R}^d$ ) then

$$(8.58) \quad \xi(A) := \sum_{i \in I} \alpha_i(A),$$

where the choice of  $\alpha_i$  in the orbit  $\tilde{\alpha}_i$  is irrelevant. We may now state:

**Proposition 8.8.** *Assume that the sequence  $(\xi_n)$  in  $\tilde{\mathcal{X}}^{(2)}$  converges to  $\xi$  w.r.t to the metric  $\mathbf{D}^{(2)}$ . Assume in addition that this sequence is diagonally tight, meaning that*

$$(8.59) \quad \lim_{M \rightarrow \infty} \sup_{n \in \mathbb{N}} \xi_n(\{(u, v) : |u - v| > M\}) = 0.$$

*Then, the sequence  $\widetilde{\text{Pr}}(\xi_n)$  converges to  $\widetilde{\text{Pr}}(\xi)$  w.r.t. to the metric  $\mathbf{D}$ .*

Let us remark that any *finite* collection of elements in  $\tilde{\mathcal{X}}^{(2)}$  is diagonally tight, since any probability measure on the Polish space  $\mathbb{R}^d$  is tight and the total mass of an element in  $\tilde{\mathcal{X}}^{(2)}$  is bounded by one.

*Proof of Proposition 8.8.* We write

$$(8.60) \quad \mathbf{D}(\widetilde{\text{Pr}}(\xi_n), \widetilde{\text{Pr}}(\xi)) \leq \mathbf{D}(\widetilde{\text{Pr}}(\xi_n), \widetilde{\text{Pr}}_A(\xi_n)) + \mathbf{D}(\widetilde{\text{Pr}}_A(\xi_n), \widetilde{\text{Pr}}_A(\xi)) + \mathbf{D}(\widetilde{\text{Pr}}_A(\xi), \widetilde{\text{Pr}}(\xi)).$$

The first term can be estimated from above by  $2 \times \xi_n(\mathcal{D}_A^c)$  by (8.27), which tends to zero as  $A$  tends to infinity, uniformly in  $n$ , from our assumption. The second term tends to zero as  $n \rightarrow \infty$  for every  $A$  by the continuity of  $\widetilde{\text{Pr}}_A$  proven in Proposition 8.1. Finally, the last term above tends to zero as  $A$  tends to infinity, since by (8.27) it can be estimated from above by  $2 \times \xi(\mathcal{D}_A^c)$ . This concludes the proof.  $\blacksquare$

#### APPENDIX A. ON THE MUKHERJEE-VARADHAN TOPOLOGY

This section collects the most important ingredients about the topology introduced by Mukherjee and Varadhan in [16, Section 3]. Similarly to Section 2 we write  $\mathcal{M}_1 = \mathcal{M}_1(\mathbb{R}^d)$  for the space of probability measures on  $\mathbb{R}^d$  and  $\mathcal{M}_{\leq 1} = \mathcal{M}_{\leq 1}(\mathbb{R}^d)$  for the space of sub-probability measures on  $\mathbb{R}^d$ . For any  $\alpha \in \mathcal{M}_{\leq 1}$  we define its orbit  $\tilde{\alpha} = \{\alpha * \delta_x : x \in \mathbb{R}^d\}$ , and we identify two measures if they have the same orbit. This introduces an equivalence relation on  $\mathcal{M}_{\leq 1}$  of which  $\widetilde{\mathcal{M}}_{\leq 1}$  denotes the corresponding quotient space. For  $k \geq 2$ , define  $\mathcal{F}_k$  as the space of continuous functions  $f: (\mathbb{R}^d)^k \mapsto \mathbb{R}$  that are *translation invariant*, i.e.

$$(A.1) \quad f(u_1 + x, \dots, u_k + x) = f(u_1, \dots, u_k), \quad \forall x, u_1, \dots, u_k \in \mathbb{R}^d,$$

and *vanishing at infinity*, in the sense that

$$(A.2) \quad \lim_{\max_{i \neq j} |u_i - u_j| \rightarrow \infty} f(u_1, \dots, u_k) = 0.$$

For  $k \geq 2$ ,  $f \in \mathcal{F}_k$  and  $\alpha \in \mathcal{M}_{\leq 1}$ , write

$$(A.3) \quad \Lambda(f, \alpha) := \int f(u_1, \dots, u_k) \prod_{1 \leq i \leq k} \alpha(du_i),$$

which only depends on  $\tilde{\alpha}$ . Define

$$(A.4) \quad \mathcal{F} := \bigcup_{k \geq 2} \mathcal{F}_k,$$

for which there exists a countable dense set (under the uniform metric) denoted by

$$(A.5) \quad \{f_r(u_1, \dots, u_{k_r}), r \in \mathbb{N}\}.$$

One can then define the following set of empty, finite, or countable collections of sub-probability measure orbits:

$$(A.6) \quad \tilde{\mathcal{X}} := \left\{ \xi = \{\tilde{\alpha}_i\}_{i \in I} : \tilde{\alpha}_i \in \tilde{\mathcal{M}}_{\leq 1}, \sum_{i \in I} \alpha_i(\mathbb{R}^d) \leq 1 \right\},$$

see Equation (16) in [16]. For every  $\xi_1, \xi_2 \in \tilde{\mathcal{X}}$ , one can then define the following metric on  $\tilde{\mathcal{X}}$ :

$$(A.7) \quad \mathbf{D}(\xi_1, \xi_2) := \sum_{r \geq 1} \frac{1}{2^r} \frac{1}{1 + \|f_r\|_\infty} \left| \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f_r, \alpha) - \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f_r, \alpha) \right|.$$

It was then shown in [16] that the space  $\tilde{\mathcal{X}}$  equipped with  $\mathbf{D}$  is a compact metric space and that  $\tilde{\mathcal{M}}_1$  is dense in  $\tilde{\mathcal{X}}$ .

#### ACKNOWLEDGEMENTS

D.E. was supported by the National Council for Scientific and Technological Development - CNPq via a Bolsa de Produtividade 303520/2019-1 and 303348/2022-4 and via a Universal Grant (Grant Number 406001/2021-9). D.E. moreover acknowledges support by the Serrapilheira Institute (Grant Number Serra-R-2011-37582). J.P. is supported by the ANR-22-CE40-0012 LOCAL. Both authors would like to thank an anonymous referee for helpful remarks on a previous version of this paper.

#### REFERENCES

- [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] J.-D. Deuschel and D. W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1989.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] M. D. Donsker and S. R. S. Varadhan. Asymptotics for the polaron. *Comm. Pure Appl. Math.*, 36(4):505–528, 1983.

- [13] D. Erhard and J. Poisat. Uniqueness and tube property for the swiss cheese large deviations, 2023.
- [14] 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.
- [15] P.-L. Lions. The concentration-compactness principle in the calculus of variations. The locally compact case, part 1. *Annales de l'I.H.P. Analyse non linéaire*, 1(2):109–145, 1984.
- [16] C. Mukherjee and S. R. S. Varadhan. Brownian occupation measures, compactness and large deviations. *Ann. Probab.*, 44(6):3934–3964, 2016.
- [17] P. Phetpradap. *Intersections of random walks*. PhD thesis, University of Bath, 2011.
- [18] M. van den Berg, E. Bolthausen, and F. den Hollander. Moderate deviations for the volume of the Wiener sausage. *Ann. of Math. (2)*, 153(2):355–406, 2001.
- [19] S. R. S. Varadhan. *Large deviations*, volume 27 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2016.
- [20] S. R. S. Varadhan. The role of topology in large deviations. *Expo. Math.*, 36(3-4):362–368, 2018.