



# 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. 2023. hal-04054395v1

**HAL Id: hal-04054395**

**<https://hal.science/hal-04054395v1>**

Preprint submitted on 31 Mar 2023 (v1), last revised 11 Apr 2024 (v2)

**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 new 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 [12]. This topology is natural for many translation-invariant problems such as the Swiss cheese model [14]. We also adapt our result to certain rescaled random walks.

## CONTENTS

1. Introduction	1
2. Topology on the space of probability measures modulo shifts	3
2.1. Vague and weak convergence	3
2.2. Widely separated sequences	4
2.3. Totally disintegrating sequences	5
2.4. Compactification	6
3. Large Deviation Principles	9
4. Lower semi-continuity of the rate function	11
5. Lower bound	13
6. Upper bound	14
7. Adaptation to rescaled random walks	16
Acknowledgements	18
References	18

## 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 [7, (II)] showed that the empirical

---

*Date:* March 31, 2023.

(D. Erhard) *Instituto de Matemática - Universidade Federal da Bahia*  
(J. Poisat) *Université Paris-Dauphine, CNRS, UMR [7534], CEREMADE, PSL Research University, 75016 Paris, France .*

measure defined by

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

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 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  $A \subset \Sigma$  such that  $\lambda(A) > 0$ , there exists  $k \in \mathbb{N}$  such that  $\pi^k(x, A) > 0$ , see [8, 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 [4, 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) [10] or folding on a large torus [1, 3, 9]. Quite recently, Mukherjee and Varadhan [12] 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. Under this new topology, they were able to prove a *strong* large deviation principle for the empirical measure of Brownian motion [12, Theorem 4.1], which was then successfully applied to the so-called polaron problem [2, 11, 12].

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

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

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 deviation for the volume of a Wiener sausage in  $\mathbb{R}^d$  ( $d \geq 2$ ), by van den Berg, Bolthausen and den Hollander [14]. In [14] (as well in the discrete random walk counterpart studied a bit later by Phetpradap [13] in his Ph.D thesis) the authors used the aforementioned folding procedure on a large torus to deal with the lack of compactness of the state space. In a future paper, we plan to 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 basically means that the empirical measure introduced in [14] is close to the minimizer of the variational problem appearing in [14].

The present paper is organized as follows. In Section 2 we introduce the relevant notations and topology. Although there is a lot in common with [12], 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 is stated in Section 3, see Theorem 3.1. 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, in which we use the well-known fact that  $(X_n, X_{n+1})$  is itself a Markov chain. Finally we adapt our result to the case of certain rescaled random walks in Section 7

We refer the reader to [4, 5, 6, 15] for an account on Large Deviation Theory and [16] for the role of topology in this theory.

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

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 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

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

For  $k \geq 1$ ,  $f \in \mathcal{F}_{2k}$  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 [12]. Indeed, in [12] 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.

**2.1. Vague and weak convergence.** In this section we give a slight generalisation of [12, Lemma 2.2] which will be useful for later purposes.

**Lemma 2.2.** 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 2.2.* The only part that was not explicitly mentioned in [12] is that  $\alpha_n$  and  $\beta_n$  can be chosen to have disjoint supports. However, this is a direct consequence of the proof in [12]. 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

$$(2.5) \quad \mu_n(B(0, R_n)) \leq \alpha(\mathbb{R}^d) + \frac{1}{R_n}.$$

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  $\mathcal{F}_4$ ,

$$(2.6) \quad \lim_{n \rightarrow \infty} \int_{(\mathbb{R}^d)^4} V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) = 0.$$

The following lemma, which mimicks [12, Lemma 2.4], lists the most important properties of widely separated sequences of sub-probability measures.

**Lemma 2.3.** *Let  $(\alpha_n)$  and  $(\beta_n)$  be two widely separated sequences in  $\mathcal{M}_{\leq 1}^{(2)}$ . Then,*

(1) *For every  $W \in \mathcal{F}_4$ ,*

$$(2.7) \quad \lim_{n \rightarrow \infty} \int W(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) = 0.$$

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

$$(2.8) \quad \lim_{n \rightarrow \infty} |\Lambda(f, \alpha_n + \beta_n) - \Lambda(f, \alpha_n) - \Lambda(f, \beta_n)| = 0.$$

*Proof of Lemma 2.3.* Let  $W \in \mathcal{F}_4$  be arbitrary and let  $V \in \mathcal{F}_4$  be the positive function from (2.6). 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  $v_1, u_2, v_2 \in \mathbb{R}^d$ ,

$$(2.9) \quad |W(0, v_1, u_2, v_2)| \leq C_\varepsilon V(0, v_1, u_2, v_2) + \varepsilon.$$

Thus

$$(2.10) \quad \begin{aligned} & \limsup_{n \rightarrow \infty} \int |W(u_1, v_1, u_2, v_2)| \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) \\ & \leq C_\varepsilon \limsup_{n \rightarrow \infty} \int V(0, v_1 - u_1, u_2 - u_1, v_2 - u_1) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) + \varepsilon \\ & = \varepsilon, \end{aligned}$$

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

$$(2.11) \quad \int f(u_1, v_1, u_2, v_2, \dots, u_k, v_k) \alpha_n(u_1, v_1) \beta_n(u_2, v_2) \prod_{3 \leq i \leq k} \gamma_{n,i}(u_i, v_i),$$

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  $v_1, u_2, v_2, \dots, u_k, v_k \in \mathbb{R}^d$ ,

$$(2.12) \quad |f(0, v_1, u_2, v_2, \dots, u_k, v_k)| \leq C_\varepsilon V(0, v_1, u_2, v_2) + \varepsilon,$$

which allows to conclude. ■

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

**Lemma 2.4.** *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 2.4.* Let  $V \in \mathcal{F}_4$  and  $\varepsilon > 0$ . By tightness of  $(\alpha_n)$  and boundedness of  $V$ , there exists  $M > 0$  such that for all  $n \geq 1$ ,

$$(2.13) \quad \int V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) \leq \int_{|u_1|, |v_1| \leq M} (\dots) + \varepsilon \|V\|_\infty.$$

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

$$(2.14) \quad \int_{\substack{|u_1|, |v_1| \leq M \\ |u_2|, |v_2| \leq 2M}} (\dots) + \int_{\substack{|u_1|, |v_1| \leq M \\ |u_2|, |v_2| > 2M}} (\dots).$$

We claim that we can make the second term smaller than  $\varepsilon$  by choosing  $M$  even larger if necessary. Indeed, since  $|u_2 - u_1| \geq M$  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  $r > 0$ ,

$$(2.15) \quad \lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} \mu_n(B((x, x), r)) = 0.$$

From now on and unless stated otherwise the balls are defined with respect to the  $\ell_\infty$ -norm. Following [12, Lemma 2.3], we obtain the following result.

**Lemma 2.5.** *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  $V \in \mathcal{F}_4$  such that  $\lim_{n \rightarrow \infty} \Lambda(V, \mu_n) = 0$ .*
- (2) *For any  $V \in \mathcal{F}_4$ ,  $\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} \int V(0, x, u, v) \mu_n(du, dv) = 0$ .*
- (3) *For any  $V \in \mathcal{F}_4$ ,  $\lim_{n \rightarrow \infty} \Lambda(V, \mu_n) = 0$ .*

*Proof of Lemma 2.5.* We prove that (1)  $\Rightarrow$  (2.15)  $\Rightarrow$  (2)  $\Rightarrow$  (3). Clearly, (3) implies (1).

(i) Let us prove that (1) implies (2.15). Letting

$$(2.16) \quad \delta := \min_{a, b, c \in B(0, 2r)} V(0, a, b, c) > 0,$$

we get

$$(2.17) \quad \begin{aligned} & \int V(u_1, v_1, u_2, v_2) \mu_n(du_1, dv_1) \mu_n(du_2, dv_2) \\ &= \int V(0, v_1 - u_1, u_2 - u_1, v_2 - u_1) \mu_n(du_1, dv_1) \mu_n(du_2, dv_2) \\ &\geq \delta \int_{v_1, u_2, v_2 \in B(u_1, 2r)} \mu_n(du_1, dv_1) \mu_n(du_2, dv_2) \\ &\geq \delta \sup_{x \in \mathbb{R}^d} \mu_n(B((x, x), r))^2. \end{aligned}$$

(ii) From (2.15) to (2). Let  $x \in \mathbb{R}^d$ . Fix  $M > 0$  and write

$$(2.18) \quad \begin{aligned} & \int |V(0, x, u, v)| \mu_n(du, dv) \\ &= \int_{B((x, x), M)} |V(0, x, u, v)| \mu_n(du, dv) + \int_{B((x, x), M)^c} |V(0, x, u, v)| \mu_n(du, dv). \end{aligned}$$

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 (2.15), 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

$$(2.19) \quad \begin{aligned} \Lambda(V, \mu_n) &\leq \sup_{x, y \in \mathbb{R}^d} \int V(0, y - x, u - x, v - x) \mu_n(du, dv) \\ &\leq \sup_{y \in \mathbb{R}^d} \int V(0, y, u, v) \mu_n(du, dv), \end{aligned}$$

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

$$(2.20) \quad \lim_{n \rightarrow \infty} \Lambda(V, \mu_n) = 0.$$

*Proof of Corollary 2.6.* Apply Item (3) in Lemma 2.5 to

$$(2.21) \quad W(u_1, v_1, u_2, v_2) := \sup_{u_3, v_3, \dots, u_k, v_k \in \mathbb{R}^d} |V(u_1, v_1, \dots, u_k, v_k)|.$$

$\blacksquare$

**2.4. Compactification.** Let us define

$$(2.22) \quad \mathcal{F}^{(2)} := \bigcup_{k \geq 2} \mathcal{F}_{2k},$$

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

$$(2.23) \quad \{f_r(u_1, v_1, \dots, u_{k_r}, v_{k_r}) : r \in \mathbb{N}\},$$

(same arguments as in [12, Section 2.2]). We define

$$(2.24) \quad \tilde{\mathcal{X}}^{(2)} := \left\{ \xi = \{\tilde{\alpha}_i\}_{i \in I} : \alpha_i \in \mathcal{M}_{\leq 1}^{(2)}, \sum_{i \in I} \alpha_i(\mathbb{R}^d \times \mathbb{R}^d) \leq 1 \right\},$$

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

$$(2.25) \quad \mathbf{D}_2(\xi_1, \xi_2) := \sum_{r \geq 1} \frac{1}{2^r} \frac{1}{1 + \|f_r\|_\infty} \left| \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f_r, \alpha) - \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f_r, \alpha) \right|.$$

**Proposition 2.7.**  $\mathbf{D}_2$  is a metric on  $\tilde{\mathcal{X}}^{(2)}$ .

*Proof of Proposition 2.7.* This follows the same line of proof as [12, Theorem 3.1].

**Step 1.** If  $\mathbf{D}_2(\xi_1, \xi_2) = 0$  then for all  $k \geq 2$  and  $f \in \mathcal{F}_{2k}$ ,

$$(2.26) \quad \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f, \alpha) = \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f, \alpha).$$

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

$$(2.27) \quad \sum_{\tilde{\alpha} \in \xi_1} \Lambda(f, \alpha)^r = \sum_{\tilde{\alpha} \in \xi_2} \Lambda(f, \alpha)^r.$$

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

$$(2.28) \quad \begin{aligned} &g_N(u_1, v_1, \dots, u_k, v_k, u_{k+1}, v_{k+1}, \dots, u_{2k}, v_{2k}) \\ &:= f(u_1, v_1, \dots, u_k, v_k) f(u_{k+1}, v_{k+1}, \dots, u_{2k}, v_{2k}) \varphi\left(\frac{u_{k+1} - u_1}{N}\right), \end{aligned}$$

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

$$(2.29) \quad \lim_{N \rightarrow \infty} \Lambda(g_N, \alpha) = \Lambda(f, \alpha)^2.$$

The general case follows the same idea.

**Step 2.** Same as in [12].

**Step 3.** This step is also an adaptation of [12], 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  $f \in \mathcal{F}_{2k}$ . Adapting [12], from these values we get those of

$$(2.30) \quad \prod_{j=1}^k \phi(s_j, t_j), \quad \text{where} \quad \phi(s, t) := \int e^{i\langle (s, t), (u, v) \rangle} \alpha(du, dv), \quad s, t \in \mathbb{R}^d,$$

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

$$(2.31) \quad \forall k \geq 1, \quad \prod_{j=1}^k \phi(s_j, t_j) = \prod_{j=1}^k \psi(s_j, t_j), \quad \text{whenever} \quad \sum(s_j + t_j) = 0.$$

Following [12], 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  $j$ . In particular,

$$(2.32) \quad \chi(s_1 + s_2, t_1 + t_2) = \chi(s_1, t_1)\chi(s_2, t_2),$$

hence as in [12] 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 2.8.** *The space  $\tilde{\mathcal{X}}^{(2)}$  equipped with  $\mathbf{D}_2$  is a compact metric space.*



*Proof of Proposition 2.8.* This proof is an adaptation of the arguments used in [12], so we only sketch the arguments.

**Step 1.** Let us first show that  $\widetilde{\mathcal{M}}_1^{(2)}$  is dense in  $\widetilde{\mathcal{X}}^{(2)}$ . Let  $\xi = \{\tilde{\alpha}_i, i \in I\} \in \widetilde{\mathcal{X}}^{(2)}$  and  $\varepsilon > 0$ . Pick a finite collection  $\{\tilde{\alpha}_i, 1 \leq i \leq k\}$  such that  $\sum_{i \geq k} p_i < \varepsilon$ , 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

$$(2.33) \quad \mu_M := \sum_{1 \leq i \leq k} \alpha_i * \delta_{(a_{i,M}, a_{i,M})} + \left(1 - \sum_{1 \leq i \leq k} p_i\right) \lambda_M \in \mathcal{M}_1^{(2)}.$$

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 (2.6) and all the mass of  $\lambda_M$  vanishes in the limit as  $M$  tends to infinity, i.e.,  $\lambda_M$  is totally disintegrating, see (2.15), we get by Lemmas 2.3 and 2.5 that for all  $k \geq 1$  and  $f \in \mathcal{F}_{2k}$ ,

$$(2.34) \quad \lim_{M \rightarrow \infty} \Lambda(f, \mu_M) = \sum_{1 \leq i \leq k} \Lambda(f, \tilde{\alpha}_i).$$

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  $\widetilde{\mathcal{M}}_1^{(2)}$ . Together with the first step this implies the result. We start with some preliminary considerations. We use the following *concentration function*:

$$(2.35) \quad q_\mu(r) := \sup_{x \in \mathbb{R}^d} \mu(B((x, x), r)), \quad r \geq 0, \quad \mu \in \mathcal{M}_1^{(2)}.$$

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,

$$(2.36) \quad q := \lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} q_{\mu_n}(r) \quad \text{and} \quad p := \lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^{2d}).$$

If  $q = 0$ , then it follows by Corollary 2.6 that  $\tilde{\mu}_n \rightarrow 0$  in  $\widetilde{\mathcal{X}}^{(2)}$ . 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

$$(2.37) \quad \lambda_n(B((0, 0), r)) \geq q/2.$$

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 2.2 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 2.4\*, for every  $V \in \mathcal{F}_4$ ,

$$(2.38) \quad \lim_{n \rightarrow \infty} \int V(u_1, v_1, u_2, v_2) \alpha_n(du_1, dv_1) \beta_n(du_2, dv_2) = 0.$$

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  $\widetilde{\mathcal{X}}^{(2)}$ , similarly to [12, Theorem 3.2]. To

---

\*The use of Lemma 2.4 corrects a small gap in [12]. Indeed, the analogous step in the proof of [12, Theorem 3.2, Step 2] does not seem to follow from Lemma 2.4, as it is claimed therein.

conclude the result we can now proceed in very much the same way as in [12]. 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  $V \in \mathcal{F}_4$  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 (2.37). 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

$$(2.39) \quad \mu_n = \sum_{j=1}^k \alpha_{n,j} * \delta_{(a_{n,j}, a_{n,j})} + \beta_n,$$

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  $V \in \mathcal{F}_4$  and  $1 \leq i < j \leq k$ ,

$$(2.40) \quad \begin{aligned} & \int V(u_1 + a_{n,i}, v_1 + a_{n,i}, u_2 + a_{n,j}, v_2 + a_{n,j}) \alpha_{n,i}(du_1, dv_1) \alpha_{n,j}(du_2, dv_2) \\ & \text{and} \quad \int V(u_1 + a_{n,i}, v_1 + a_{n,i}, u_2, v_2) \alpha_{n,i}(du_1, dv_1) \beta_n(du_2, dv_2) \end{aligned}$$

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 [12, Theorem 3.2] for details. ■

### 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 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 set  $A \subseteq \mathbb{R}^d$ .
- (2) **(Irreducibility)** For all Borel set  $A \subseteq \mathbb{R}^d$  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

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

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 [8, Corollary 3.4 and Equation (4.1)]. Assumption (3) is used during the proof of Lemma 6.2.

Let us denote by

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

the pair empirical measure associated to  $X$  and  $\tilde{L}_n$  be its orbit in  $\widetilde{\mathcal{M}}_1^{(2)}$ . We obtain the following:

**Theorem 3.1.** *As  $n \rightarrow \infty$ ,  $(\tilde{L}_n)_{n \in \mathbb{N}}$  satisfies a strong Large Deviation Principle on the compact metric space  $\widetilde{\mathcal{X}}^{(2)}$  equipped with  $\mathbf{D}_2$ , with speed  $n$  and rate function  $\tilde{J}^{(2)}$ , where*

$$(3.3) \quad \tilde{J}^{(2)}(\xi) := \sum_{i \in I} h(\alpha_i | \alpha_{i,1} \otimes \pi), \quad (\xi = \{\tilde{\alpha}_i\}_{i \in I}),$$

$\alpha_i$  being any representative of  $\tilde{\alpha}_i$  and  $\alpha_{i,1}$  its projection onto the first  $d$  coordinates. Here,  $h(\cdot | \cdot)$  is the relative entropy between two measures, i.e.,

$$(3.4) \quad h(\mu | \nu) = \int \log \left( \frac{d\mu}{d\nu} \right) d\mu,$$

with the understanding that the integral is infinite if the Radon Nikodym derivative above does not exist.

**Remark 3.2.** *We note that the rate function in (3.3) 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 (3.3) the choice of the element  $\alpha_i$  in the orbit  $\tilde{\alpha}_i$  does not matter.*

Let us recall 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

$$(3.5) \quad \pi^{(2)}(x_1, x_2, dy_1, dy_2) := \delta_{x_2}(y_1) \pi(x_2, dy_2), \quad x_1, x_2, y_1, y_2 \in \mathbb{R}^d.$$

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 end this section with a corollary. To that end define the empirical measure  $L_n^{(1)}$  of  $(X_i)_{i \in \mathbb{N}_0}$  via

$$(3.6) \quad L_n^{(1)} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

Note that  $L_n^{(1)}$  is simply the second marginal of  $L_n$ . We further denote by  $\tilde{L}_n^{(1)}$  the orbit of  $L_n^{(1)}$  in  $\widetilde{\mathcal{M}}_1(\mathbb{R}^d)$ , the space defined in [12, Section 2]. See Remark 2.1 for the difference

between  $\widetilde{\mathcal{M}}_1^{(2)}$  and  $\widetilde{\mathcal{M}}_1(\mathbb{R}^{2d})$ . We moreover denote by  $\widetilde{\mathcal{X}}$  the compactification of  $\widetilde{\mathcal{M}}_1(\mathbb{R}^d)$ , equipped with the metric  $\mathbf{D}$ , see [12, Section 3] for details. The following corollary is then a direct consequence of Theorem 3.1 and the contraction principle.

**Corollary 3.3.** *As  $n \rightarrow \infty$ ,  $(\widetilde{L}_n^{(1)})_{n \in \mathbb{N}}$  satisfies a strong Large Deviation Principle on the compact metric space  $\widetilde{\mathcal{X}}$  equipped with  $\mathbf{D}$ , with speed  $n$  and rate function  $\widetilde{J}^{(1)}$  defined by*

$$(3.7) \quad \widetilde{J}^{(1)}(\theta) = \inf\{\widetilde{J}^{(2)}(\xi) : \xi_1 = \theta\},$$

where for  $\xi = \{\widetilde{\alpha}_i\}_{i \in I} \in \widetilde{\mathcal{X}}^{(2)}$  we denote by  $\xi_1$  the collection  $\{\alpha_{i,2}\}_{i \in I}$  and  $\alpha_{i,2}$  denotes the second marginal of  $\alpha_i$  (projection onto the  $d$  last coordinates).

*Proof of Corollary 3.3.* In order to apply the contraction principle [4, Theorem 4.2.1], we show that the second projection (onto the last  $d$  coordinates) from  $(\widetilde{\mathcal{X}}^{(2)}, \mathbf{D}_2)$  to  $(\widetilde{\mathcal{X}}, \mathbf{D})$  is continuous. The reader may check that it is actually enough to prove continuity of the projection from  $(\widetilde{\mathcal{M}}_{\leq 1}^{(2)}, \mathbf{D}_2)$  to  $(\widetilde{\mathcal{M}}_{\leq 1}, \mathbf{D})$ . Suppose that  $(\widetilde{\nu}_n(du, dv))$  is a sequence of orbits of sub-probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$  that converges to  $\widetilde{\nu}(du, dv)$  for the  $\mathbf{D}_2$  metric. Let  $k \geq 2$  and  $f \in \mathcal{F}_k$ . We may write  $f(v_1, \dots, v_k) = \hat{f}(u_1, v_1, \dots, u_k, v_k)$  where  $\hat{f} \in \mathcal{F}_{2k}$  is constant along the  $u$ -variables. Then (the choice of elements in the orbits below is not relevant),

$$(3.8) \quad \begin{aligned} \int f(v_1, \dots, v_k) \prod_{1 \leq i \leq k} \nu(\mathbb{R}^d, dv_i) &= \int \hat{f}(u_1, v_1, \dots, u_k, v_k) \prod_{1 \leq i \leq k} \nu(du_i, dv_i) \\ &= \lim_{n \rightarrow \infty} \int \hat{f}(u_1, v_1, \dots, u_k, v_k) \prod_{1 \leq i \leq k} \nu_n(du_i, dv_i) \\ &= \lim_{n \rightarrow \infty} \int f(v_1, \dots, v_k) \prod_{1 \leq i \leq k} \nu_n(\mathbb{R}^d, dv_i). \end{aligned}$$

This concludes the proof. ■

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

**Proposition 4.1.** *The function*

$$(4.1) \quad \xi = \{\widetilde{\alpha}_i\}_{i \in I} \in \widetilde{\mathcal{X}}^{(2)} \mapsto \widetilde{J}^{(2)}(\xi) = \sum_{i \in I} h(\alpha_i | \alpha_{i,1} \otimes \pi)$$

*is lower semi-continuous.*

*Proof of Proposition 4.1.* We use the notation  $J^{(2)}(\alpha) := h(\alpha | \alpha_1 \otimes \pi)$ , where  $\alpha \in \mathcal{M}_{\leq 1}^{(2)}$  and  $\alpha_1$  stands for its projection on the  $d$  first coordinates along the proof. Recall (3.5). 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  $u: \mathbb{R}^d \times \mathbb{R}^d \rightarrow [1, +\infty)$ , see [4, Theorem 6.5.12 and Corollary 6.5.10]. Let  $(\mu_n)$  be a sequence in  $\widetilde{\mathcal{X}}^{(2)}$  converging to  $\xi = \{\widetilde{\alpha}_i\}_{i \in I}$ . 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 2.8 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  $\mathbb{R}^d \times \mathbb{R}^d$ , and  $a_n^{(i)}$  ( $1 \leq i \leq k$ ) are sequences in  $\mathbb{R}^d$ . Moreover those sequences can be chosen such that

$$(4.4) \quad \begin{aligned} \alpha_n^{(i)} * \delta_{(a_n^{(i)}, a_n^{(i)})} &\Rightarrow \alpha_i \in \tilde{\alpha}_i, \quad n \rightarrow \infty, \\ \lim_{n \rightarrow \infty} \min_{i \neq j} |a_n^{(i)} - a_n^{(j)}| &= +\infty, \end{aligned}$$

and  $(\beta_n)$  is widely separated from each  $(\alpha_n^{(i)})$ , with (recall (2.35))

$$(4.5) \quad \limsup_{n \rightarrow \infty} q_{\beta_n}(r) \leq \varepsilon, \quad \forall r > 0.$$

Moreover, by Lemma 2.2 and the construction in Step 2 of Proposition 2.8 the supports of  $\alpha_n^{(1)}, \alpha_n^{(2)}, \dots, \beta_n$  are all disjoint and for each  $i$  there exists a sequence  $(R_n^{(i)})_{n \in \mathbb{N}_0}$  tending to infinity such that  $\text{Supp}(\alpha_n^{(i)}) \subseteq B((-a_n^{(i)}, -a_n^{(i)}), R_n^{(i)})$  and the support of  $\beta_n$  is contained in the complement of  $\bigcup_i B((-a_n^{(i)}, -a_n^{(i)}), R_n^{(i)})$ . We now define

$$(4.6) \quad u_i^{(n)}(x, y) = u_i(x + a_n^{(i)}, y + a_n^{(i)}), \quad x, y \in \mathbb{R}^d.$$

Note that by the compactness of the support of  $u_i$  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 (3.5) there is  $R > 0$  such that for all  $1 \leq i \leq k$

$$(4.7) \quad \begin{aligned} \text{Supp } u_i^{(n)} &\subseteq B((-a_n^{(i)}, -a_n^{(i)}), R), \\ \text{Supp } \pi^{(2)} u_i^{(n)} &\subseteq \mathbb{R}^d \times B(-a_n^{(i)}, R), \end{aligned}$$

hence

$$(4.8) \quad \left( \frac{1 + u_i^{(n)}}{1 + \pi^{(2)} u_i^{(n)}} \right)(x, y) \neq 1 \Rightarrow y \in B(-a_n^{(i)}, R).$$

Therefore

$$(4.9) \quad \lim_{n \rightarrow \infty} \int \log \left( \frac{1 + u_i^{(n)}}{1 + \pi^{(2)} u_i^{(n)}} \right) d\alpha_n^{(j)} = \begin{cases} \int \log \left( \frac{1 + u_i}{1 + \pi^{(2)} u_i} \right) d\alpha_i & (i = j) \\ 0 & (i \neq j). \end{cases}$$

Finally, letting

$$(4.10) \quad u^{(n)} = \sum_{1 \leq i \leq k} u_i^{(n)},$$

we obtain:

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

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 (4.8) 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 [12, proof of Lemma 4.2]. ■

## 5. LOWER BOUND

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

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

*Proof of Proposition 5.1.* Let  $\xi = \{\alpha_i, i \in I\}$  be an element of  $\tilde{\mathcal{X}}^{(2)}$  such that  $J^{(2)}(\xi) < +\infty$ , 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

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

We proceed as in [12, Lemma 4.3] and use the density of  $\tilde{\mathcal{M}}_1^{(2)}$  in  $\tilde{\mathcal{X}}^{(2)}$ , already proven in the proof of Proposition 2.8 (Step 1). Let  $k \geq 1$  and consider the sequence  $(\mu_M)$  defined as in (2.33), except that we replace the totally disintegrating sequence  $(\lambda_M)$  by  $(\lambda_{M,1} \otimes \pi)$ , that is totally disintegrating as well and is such that  $\tilde{\mu}_M$  still converges in  $\tilde{\mathcal{X}}^{(2)}$  to  $\xi$ . Using the sub-additive property of  $\nu \in \mathcal{M}_{\leq 1}^{(2)} \mapsto h(\nu | \nu_1 \otimes \pi)$  (straightforward from (4.2)), 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  $\tilde{\mathcal{M}}_1^{(2)}$  which converges in  $\tilde{\mathcal{X}}^{(2)}$  to  $\xi$  and is such that

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

The lower bound now follows from the standard Large Deviation lower bound of the pair empirical measure on  $\mathcal{M}_1^{(2)}$  (see [8] and the discussion in Section 1). ■

## 6. UPPER BOUND

In this section we prove the following

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

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

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  $\tilde{F}$  defined on  $\tilde{\mathcal{M}}_1^{(2)}$ . 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*

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

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

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

where

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

**Remark 6.4.** Lemma 6.3 is analogous to Lemma 4.6 in [12]. 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 [12], 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

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

so that

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

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 [12, proof of Lemma 4.5] one may then conclude, so we omit the details.  $\blacksquare$

*Proof of Lemma 6.3.* As it can be seen from the second step of the proof of Proposition 2.8, 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 (2.39), 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 [12, 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 [12, 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 unbounded 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 [12, 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 [4, 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 [12] 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  $a_n^2 = o(n)$ . Define  $\ell := \ell(\varepsilon, n) = \lfloor \varepsilon a_n^2 \rfloor$ ,  $M := \lfloor n/\ell \rfloor$  and

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

Denote by  $L_n^\varepsilon$  the corresponding pair empirical measure, that is

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

**Remark 7.1.** *The relevant scale for potential applications to the Swiss cheese model [13, 14] corresponds to the choice  $a_n := n^{1/d}$ .*

We prove the following:

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

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

Here,

$$(7.4) \quad \tilde{J}_\varepsilon^{(2)}(\xi) := \sum_{i \in I} h(\alpha_i | \alpha_{i,1} \otimes \pi_\varepsilon), \quad (\xi = \{\tilde{\alpha}_i\}_{i \in I}),$$

where  $\pi_\varepsilon$  denotes 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

$$(7.5) \quad \pi_{n,\varepsilon}^{(2)} g(x, y) = \mathbb{E} \left[ g(y, y + X_1^{\varepsilon,n}) \right], \quad \text{and} \quad \pi_\varepsilon^{(2)} g(x, y) = \mathbb{E} \left[ g(y, y + B_\varepsilon) \right].$$

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

$$(7.6) \quad \limsup_{n \rightarrow \infty} \frac{1}{M} \log \mathbb{E} \exp(M \tilde{F}(u, c, R, \tilde{L}_n^\varepsilon)) \leq 0.$$

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

$$(7.7) \quad \prod_{i=1}^M \frac{g(X_{i-1}^{\varepsilon,n}, X_i^{\varepsilon,n})}{\pi_\varepsilon^{(2)} g(X_{i-1}^{\varepsilon,n}, X_i^{\varepsilon,n})} = \prod_{i=1}^M \frac{g(X_{i-1}^{\varepsilon,n}, X_i^{\varepsilon,n})}{\pi_{n,\varepsilon}^{(2)} g(X_{i-1}^{\varepsilon,n}, X_i^{\varepsilon,n})} \prod_{i=1}^M \frac{\pi_{n,\varepsilon}^{(2)} g(X_{i-1}^{\varepsilon,n}, X_i^{\varepsilon,n})}{\pi_\varepsilon^{(2)} g(X_{i-1}^{\varepsilon,n}, X_i^{\varepsilon,n})}.$$

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

$$(7.8) \quad \lim_{n \rightarrow \infty} \sup_{x, y \in \mathbb{R}^d} \left| \frac{\pi_{n,\varepsilon}^{(2)} g(x, y)}{\pi_\varepsilon^{(2)} g(x, y)} - 1 \right| = 0.$$

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 (7.8) allows to control the rightmost factor in (7.7). The first factor on the right hand side can be dealt with as in the proof of Lemma 6.2. ■

## 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.

## REFERENCES

- [1] Erwin Bolthausen. Localization of a two-dimensional random walk with an attractive path interaction. *Ann. Probab.*, 22(2):875–918, 1994.
- [2] Erwin Bolthausen, Wolfgang König, and Chiranjib 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.
- [3] Erwin Bolthausen and Uwe Schmock. On self-attracting  $d$ -dimensional random walks. *Ann. Probab.*, 25(2):531–572, 1997.
- [4] Amir Dembo and Ofer 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.
- [5] Frank den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.
- [6] Jean-Dominique Deuschel and Daniel W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1989.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] M. D. Donsker and S. R. S. Varadhan. Asymptotics for the polaron. *Comm. Pure Appl. Math.*, 36(4):505–528, 1983.
- [11] 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.
- [12] Chiranjib Mukherjee and S. R. S. Varadhan. Brownian occupation measures, compactness and large deviations. *Ann. Probab.*, 44(6):3934–3964, 2016.
- [13] Parkpoom Phetpradap. *Intersections of random walks*. PhD thesis, University of Bath, 2011.
- [14] 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.
- [15] 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.
- [16] S. R. S. Varadhan. The role of topology in large deviations. *Expo. Math.*, 36(3-4):362–368, 2018.